A conflict early warning system for the Sahel region

Gabriele Cirotto
s307732@studenti.polito.it
Politecnico di Torino

Alberto Foresti
s309212@studenti.polito.it
Politecnico di Torino

Luca Varriale
s300795@studenti.polito.it
Politecnico di Torino

ABSTRACT

This report addresses the escalating threats to peace and development in the Sahel region due to political and social conflicts. It introduces an experimental initiative for a comprehensive early warning system, leveraging historical data and aligning with Sustainable Development Goals. The project involves in dataset creation, causal data augmentation, and model deployment, with a focus on predicting fatalities at country level. The use of Machine Learning models like XGBoost, and causal data augmentation aims to enhance model robustness, emphasizing the significance of calibration for confident decision-making in preventing catastrophic events.The report emphasizes caution in deploying predictive models for social conflicts, highlighting the need for data minimization and thoughtful consideration of long-term societal consequences. The associated code and dataset, available for reference and collaboration, can be accessed through the provided repository at GitHub.

1 INTRODUCTION

In recent years, peace and development have faced escalating threats due to rising internal and cross-border security challenges. These challenges encompass armed conflicts, severe terrorist attacks carried out by jihadist groups like Boko Haram, Al Qaeda, or IS-affiliated factions, as well as activities by organized crime. As a result, this leads to substantial macroeconomic and fiscal burdens. For instance, Mali nearly quadrupled its expenditure on defense, increasing from USD 132 million to USD 495 million between 2013 and 2018, as reported by the Stockholm International Peace Research Institute [31]. In the face of escalating political instability and violence in the Sahel region, there is a pressing need for an effective conflict early warning system to mitigate the impact on vulnerable populations. This report introduces a pioneering initiative aimed at developing a comprehensive early warning system for the Sahel, designed to forecast political violence and enhance public awareness. Covering twelve countries, including Senegal, Chad, Mali, and Nigeria (represented in Fig. 1), this initiative aims to explore the potential of real-time information in locations where traditional data sources are scarce. Aligned with Sustainable Development Goals 10 and 16, focusing on Reduced Inequality and Peace, Justice, and Strong Institutions, respectively, this endeavor also aligns with the vision set forth in the UN Secretary General's Data Strategy, emphasizing the imperative of harnessing data for more effective decision-making.

1.1 Project overview

In order to develop a robust and general model for the prediction of a state of conflict, we carried out three main steps:

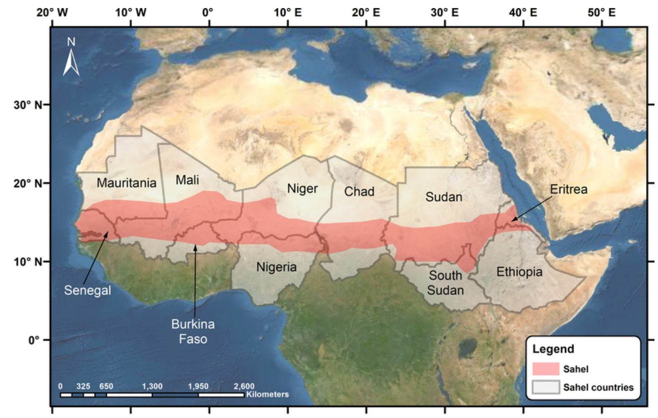


Figure 1: The red stripe represents the Sahel region [3].

- **Dataset Creation:** we constructed a dataset with significant features for the task, coming from well-known and publicly available datasets.
- **Causal Data Augmentation:** in order to train a more robust model, we exploited a mechanism to generate synthetic data, employing state-of-the-art methods in this field.
- **Model Deployment:** for assessing the quality of the dataset and its augmented version, we employed different machine learning models as XGBoost, which have proven to be efficient models for time series forecasting [10].

2 RELATED WORKS

2.1 Indicators of Conflict

Bintu Zahara Sakor [25] analyzes the correlation between population growth, peace, and structural security, illustrating how a youthful demographic structure, or large youth cohorts, can potentially pose a threat to national security. Climate-related data were not taken into account. As indicated by [20] and [26], predictions of conflicts or asylum migration in Africa do not show significant improvement when climate changes are considered. In our work, we utilized various conflict indicators, which were thoroughly analyzed in detail in Section 4.1.

2.2 Time Series Forecasting

Time Series Forecasting serves as a pivotal tool in predictive analysis, enabling organizations to anticipate future trends and make informed decisions based on historical data patterns. One prominent application of this technique is exemplified through its integration into various domains for proactive planning [15], resource allocation [8], and risk mitigation [12]. The effectiveness of time series forecasting lies in its ability to harness historical data, enabling organizations to make timely and well-informed decisions across various domains. However, the value of time series forecasting hinges on prompt and appropriate responses to the insights provided. In our work we focused on the state-of-the-art models in time series forecasting, as XGBoost and LightGBM.

2.3 African Early Warning Systems

ECOWARN is a monitoring tool implemented in 2003 as a mechanism for conflict prevention, management, resolution, peacekeeping and security for countries in West Africa. Different assessments, like [23] and [1] concludes that, despite the early warning network effectively gathering collaborative intelligence on transnational crimes throughout the sub-region, member states of ECOWAS have regrettably fallen short in responding promptly to these identified security threats within their respective jurisdictions. The research underscores a fundamental principle: without a timely response, the efficacy of early warning systems is jeopardized, rendering them essentially worthless in achieving their intended objectives.

Another important data-driven conflict early-warning tool is the ViEWS system [14], developed for Africa and the Middle East and highlighted by United Nations. It provides estimates of armed conflict probabilities at country and sub-national levels for the next 1–36 months. The system's monthly updated forecasts, accessible through an API and ViEWS website, contribute to proactive peace initiatives and addressing uncertainties, and supporting UNISS (United Nations Integrated Strategy for the Sahel) countries in the Sahel region.

2.4 Causality

Causal inference is the field of machine learning which aims at answer causal queries from data with minimum experimentation and minimum number of assumptions [22]. This allows us experimenting with data and asking questions like "*what if this country was at war in this time frame?*". This experimentation, known as causal intervention, is crucial for policy makers, as it determines the consequences of their interventions. Graphical models stand out as the main tool to execute causal queries, they can compactly capture the relationship between features, potentially explaining them in terms of mathematical functions[19]. Causal discovery is the process of computing these graphs from data, it represents a vibrant area of research, as it intersects with deep learning in disentangled representation learning [38]. LiNGAM[28] is a causal discovery method that assumes non-Gaussian exogenous factors to extract causal graphs. Over the years this model has been extended to cover multiple use cases and supports several software implementations [5].

2.5 Generalization

Machine learning (ML) methods heavily rely on large quantities of good quality data. However, in real scenarios data often lack in quantity and quality. Many datasets, for example, are not able to cover the whole data distribution over which the ML model operates. In this case, we say that the ML model does not generalize well. In this regard, foundational models, being trained on huge quantities of data have been able to partially solve the generalization problem. An example is SAM [18], a foundational model for computer vision tasks released by META in 2023 which set a new benchmark for the generalization of ML models to unseen tasks. Likewise, in Natural Language Processing (NLP), we can resort to even a wider pool of foundational models, such as GPT [7], Mistral[17] or Llama [34]. In the time-series domain, however, despite some attempts [13], foundational models still have to make a breakthrough. However, this has roots in the ineffectiveness of DL methods, including transformers, in the time series domain [29]. In addition, since state of the art time series algorithms, such as XGBoost, are relatively quick to train from scratch, it might be worth considering the generalization problem from another perspective, acting on data generation. The process of creating new data from existing data and prior knowledge is called data augmentation. The literature offers different modalities in the context of time series [21][9]. Additionally, [16] presents a connection between data augmentation and causal interventions, showing the successfulness of data augmentation techniques in domain generalization. Several works [36][4] build upon this intuition and define transformations and sampling techniques that make use of causality. In our work we attempt at gathering intuitions and techniques from the literature to develop a methodology for an end-to-end data augmentation strategy employable in a real world scenario such as the development of an early warning system for the Sahel region.

2.6 Calibration

Calibration is particularly important in applications where decision-making relies on confidence levels associated with predictions. In fields such as medical diagnosis or fraud detection, understanding the certainty of a model's prediction is essential for users to make informed choices. Poorly calibrated models may lead to misguided decisions, as predicted probabilities may not accurately reflect the true likelihood of events. Conformal prediction [27] is considered the state of the art for calibration in a predictive setting. It relies on the exchangeability assumption, which, informally, implies that, given a set of N outcomes, any of the ordering is equally likely. Under this assumption, conformal prediction makes use of a calibration set to generate prediction intervals with true theoretical guarantees. However, the exchangeability assumption often does not hold in the time series case due to the fact that the order of observation matters. Moreover, exchangeability does not account for covariate shift, a common situation in many real world applications. [33] proposes to use a likelihood ratio which relaxes the exchangeability assumption by weighing samples to make the calibration set distribution match a target distribution. This method, however require some knowledge of the likelihood ratio, which might be difficult to attain. In our work we explore data augmentation techniques in calibration.

3 METHODS

3.1 Kernel Density Estimation

In order to perform Causal Data Augmentation, we need to simulate probability distributions. However we do not know the closed form of the probability distribution we want to simulate. In this regard, non-parametric methods allow us to perform the simulation. We chose Kernel Density Estimation (KDE): given n independent identically distributed (iid) samples $X = \{x_1, ..., x_n\}$ drawn from some distribution with an unknown density f at any given point x. We estimate the shape of f:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

Where:

- K is a non-negative function. We use the Gaussian kernel throughout our work.
- h is a smoothing parameter called bandwidth. We use the Silverman bandwidth throughout our work [30].

In order to sample using KDE, we need to:

(1) Draw a value from the kernel density.
(2) Independently select one of the data points at random and add its value to the result of (1).

The proof of this result can be found at [6]. In our algorithm, however, we also need to sample from conditional distributions. Formally, we need to sample from $P(Y|X = x)$ for some random variables X, Y. Therefore, we first fit a joint kernel density to $P(X, Y)$. Then, given n iid samples $\{y_1, ..., y_n\}$, we compute a vector of probabilities $\mathcal{P} = [P(x, y_1), ..., P(x, y_n)]^T$ and we get the normalised version $\mathcal{P}'$ such that $\mathcal{P}' = \mathbf{1} \cdot \mathcal{P} = 1$, where $\mathbf{1}$ denotes a vector of all ones. We modify the previous sampling procedure using $\mathcal{P}'$:

(1) Draw a value from the conditional kernel density, calculating the Silverman bandwidth by weighing the data points $\{y_1, ..., y_n\}$ using $\mathcal{P}'$.
(2) Independently select one of the data points sampling according to $\mathcal{P}'$ and add its value to the result of (1).

We remark that, in point (1), the Silverman bandwidth determines the variance of the Gaussian kernel, which is enough for us to sample from the kernel density.

3.2 Causal Data Augmentation

In this section we describe the mechanism that we used to generate synthetic data to train models which are more robust to domain shifts. We start by laying out the assumptions which provide the foundation of our methodology. We now denote our dataset as $X \in \mathbb{R}^{m,n}$, where m is the number of samples and n is the number of predictors. Our first assumption is that, given a set of shift features Σ, we can describe the joint distribution of our dataset in terms of Σ, a set of features D that depend on Σ and a set of features I, which does not depend on Σ. We can express this dependency with the following causal graph:

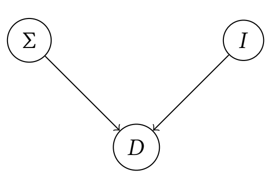


Figure 2: The primitive causal graph that we assume for our algorithm.

Additionally, as it can be seen from Fig. 2, we assume no hidden confounders between the variables since by definition they cannot be expressed in terms of the features of our dataset. However, their effect exists and we can reduce it by using powerful predictors.

In order to understand the causal relations between variables and their dependence, we employed the LiNGAM causal discovery method [28] followed by a post-processing step where we removed the edges not conform to the structure we presented in Fig. 2. We computed the causal graph using data from all the countries, assuming that all countries undergo similar dynamics in terms of conflicts. Using the causal graph of Fig. 2, we can write the joint distribution of our dataset as:

$$f_{\Sigma,I,D}(\sigma, i, d) = f_{\Sigma}(\sigma) f_I(i) f_{D|\Sigma,I}(d|\sigma, i) \tag{2}$$

Using this description, we split the dataset X in two datasets Y, Z, representing two different domains according to some rule which employs the set of features Σ. In this way, Σ acts as representative of the domain shift. We must note that it is not always straightforward to assign samples to domains with a rule over the features, as domain shifts are not well defined. In our case, we used a political violence indicator well suited for the task.

After defining the two domains, we split the dataset X according to country data, getting datasets for every country. We denote single instances of these datasets by X_c. Then we apply the causal data augmentation algorithm described in Algorithm 1.

Algorithm 1 Causal Data Augmentation

---

procedure CAUSAL DATA AUGMENTATION($S, T, G_\Sigma, \Sigma, n$)
  $X \leftarrow G_\Sigma$ *topological sort of features in the causal graph*
  $D, I \leftarrow G_\Sigma$ *Get dependent and independent variables from causal graph*
  $F \leftarrow []$ *list of features fitting the kernel*
  $A \leftarrow []$ *initialise augmented dataset*
  for each feature f in X **do**
    $F \leftarrow f$
    if $f \in I$ **then**
      $k \leftarrow$ fit kernel on $S[F]$
      $v \leftarrow S[f]$
    else
      $k \leftarrow$ fit kernel on $T[F]$
      $v \leftarrow T[f]$
    end if
    $V \leftarrow []$
    for each sample s in A **do**
      $l \leftarrow \text{len}(v)$
      $X_s \leftarrow s[F \setminus \{f\}]$ repeated l times
      $X_s \leftarrow X_s, v$ *append v to* X_s
      $\mathcal{P}' \leftarrow k(X_s)$ *get the value of the pdf fitted by k*
      $b \leftarrow$ Silverman bandwidth calculated using $\mathcal{P}'$
      $V \leftarrow \text{sample}(v, \mathcal{P}', n) + N(\mu = 0, \sigma = b)$
    end for
    $A \leftarrow V$
  end for
  return A
end procedure

We highlight the crucial steps of the algorithm. We sample the values of one feature at a time, following the topological order of the causal graph G_Σ. In practice, given a feature f, we sample its

values following the procedure used by [4]: we sample from the probability distribution $P(f|pa(f))$, where $pa(f)$ are the causal parents of f and the predecessors of f in G_Σ. Importantly, we simulate the several conditional distributions by fitting a kernel on country data if $f \in I$, otherwise we fit on the target domain dataset. In this way we keep the characteristics of the country, while also simulating a different scenario.
In practice, we divide the dataset X in two domains, one simulating a scenario of stability, Y, and one simulating a scenario of instability Z. Additionally, we divide the dataset again to get country data X_c. In the augmentation procedure we perform both the shifts $X_c \rightarrow Y$ and $X_c \rightarrow Z$, in this way we do not need to assume the domain of X_c.

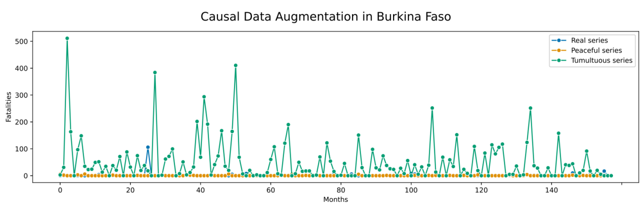


Figure 3: Fatality data augmentation in Burkina Faso.

3.3 Calibration

We employ conformal prediction to calibrate our models and produce calibration intervals. However, we noticed that conformal prediction did not perform ideally under distribution shift for similar reasons as cited in the previous section. In order to mitigate this phenomenon, we apply Causal Data Augmentation also on the calibration sets.

3.4 Metrics

We employ the Mean Absolute Error (MAE) as the reference metric for training our models and tuning the hyperparameters. However, we feel that in this context we need to add more context to our experiments. For this reason we introduce *Spike Precision* (SP) and *Spike Recall* (SR). SP and SR are defined by two parameters:

- The *Spike Threshold* α: it is the lower bound for a fatality data point to be defined a spike.
- The *Spike Tolerance* β: it defines the tolerance of the model in detecting spikes. That is, if y is the ground truth and y' is the prediction of our model, then we say that the model classified the peak correctly if:

$$|y' - y| < \beta$$

In SR we measure, among the true spikes, how many the model was able to identify correctly. While, with SP, we measure how many, among the data points classified as spikes by the model, how many were truly spikes.

4 EXPERIMENT

4.1 Dataset Creation

In our pursuit of the most pertinent features for our model, we mainly considered the following datasets:

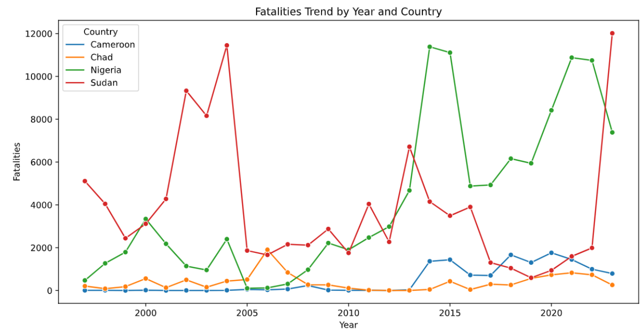


Figure 4: Fatalities trend by year.

- **ACLED** [32]: A comprehensive collection of data spotlighting humanitarian crises, including statistics on the number of victims and other pertinent details.
- **UNDP** [35]: A dataset comprising crucial dimensions of human development. The indices in this dataset and in the following are annual; consequently, we employed the interpolation technique to depict gradual monthly variations.
- **WDI** [37]: World Development Indicators - the primary World Bank compilation of development indicators, meticulously curated from internationally recognized sources.

The dataset's temporal granularity is set at a monthly cadence. Given the contextual framework of this project and its country-specific breakdown, the proposition of daily predictions is deemed impractical, as it would not contribute to practical usability.

The features extracted encompass:

- **Fatalities [2]**: This serves as the target variable underpinning our predictive efforts. In Fig. 4, the trend of fatalities per country for four Sahel countries is shown.
- **Type of Conflict [2]**: Since our task involves the prediction of fatalities , it is crucial to obtain information about the number of relevant events that have occurred, such as political violence, demonstrations, or strategic developments. In particular, political violence, as defined by ACLED, refers to the use of force by a group with a political motivation, leading to distinct political effects. Demonstrations, instead, are defined as an in-person public gathering of three or more people advocating for a shared cause. Finally, strategic developments encompass various events like recruitment drives, peace talks, and arrests.
- **HDI (Human Development Index) [24]**: A coincise measure evaluating sustained advancement across three fundamental dimensions of human development, ensuring a lengthy and healthy life, access to knowledge, and the provision of a decent standard of living. The idea behind its conception was a shift in the development discourse, steering away from an exclusive focus on economics toward a more holistic concept of 'human development'. In Fig. 5, the trend of the HDI for some countries is shown.
- **Voice and Accountability [37]:** Reflects perceptions of citizens' ability to participate in selecting their government.

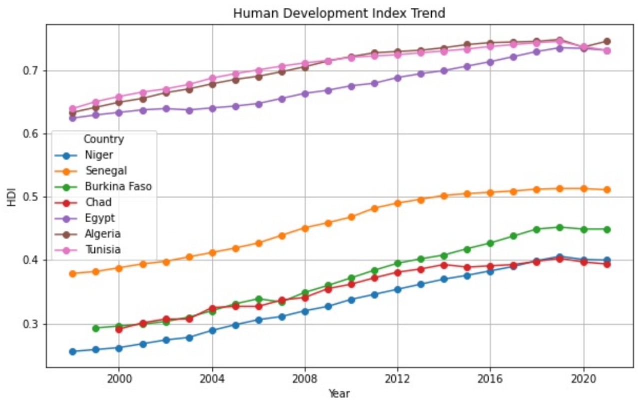


Figure 5: Comparison of HDI for some Sahel and non-Sahel African countries; the closer the value is to 1, the more the country is in a good state of human development.

- **Regulatory Quality [37]:** Reflects government's efficacy in formulating and implementing policies, indicating the effectiveness of regulatory frameworks.
- **Control of Corruption [37]:** Reflects perceptions of the extent to which public power is used for private gain, covering both petty and grand forms of corruption.
- **Political Stability and Absence of Violence/Terrorism [37]:** Measures the likelihood of political instability and violence, providing insights into a country's overall political stability.
- **Rule of Law [37]:** Reflects perceptions of societal rule adherence, encompassing contract enforcement, property rights, policing, courts, and crime likelihood.

4.2 Experiment Configuration

We mainly experiment with predicting fatalities with a one month warning, however we have found similar results across other time lags. In order to test our pipeline, we employed three different models: XGBoost, LightGBM and Random Forest Regressor. We split the dataset into three datasets for training, validation and testing. The training dataset was composed of the 60% of the entries in the original dataset, while validation and testing accounted for 20% each. In order to train our models we performed a grid search over their parameters as shown in Tab. 1. We tuned the CDA hyperparameters manually, however we focused on two split variables: *Fatalities* and *Political Stability and Absence of Violence/Terrorism*, trying different values for their split. We note that, in principle, one could design different augmentations for different countries, in the same way one treats the hyperparameters for the models. This can greatly impact the performances country-wise.

4.3 Evaluation

With the configuration illustrated in the previous section, we got the following results. For SP and SR we set $\alpha = 50$ and $\beta = 20$:

Model	**Parameter**	**Values**
XGBoost	n_estimators	50, 100, 200
	learning_rate	0.01, 0.1, 0.2
	max_depth	3, 5, 7
RandomForest	n_estimators	50, 100, 200
	max_depth	None, 5, 10
	min_samples_split	2, 5, 10
LightGBM	n_estimators	50, 100, 200
	learning_rate	0.01, 0.1, 0.2
	max_depth	-1, 5, 10

Table 1: Hyperparameters for different models.

Data Type	Model	MAE	Spike precision	Spike recall
Original	LightGBM	92.82	0.33	**0.55**
Original	RandomForest	116.47	0.33	0.40
Original	XGBoost	103.83	0.34	0.40
Augmented	LightGBM	79.75 (-14%)	**0.37**	0.44
Augmented	RandomForest	104.48 (-10%)	**0.37**	0.43
Augmented	XGBoost	**76.51 (-26%)**	0.35	0.42

As we can see the augmentation positively affects all models. Moreover, we can see how CDA improves also spike precision and spike recall in all cases, except for LightGBM: in that case spike recall is much better for the model trained with just original data. We also show the fraction of countries which improve the quality of their model with CDA with respect to MAE:

Model	Augmentation Improvement
LightGBM	0.750
RandomForest	0.583
XGBoost	0.750

In this case we can observe how CDA improves the performance for the majority of countries, especially with XGBoost and LightGBM.

4.3.1 Country Specificity. Ideally, one should choose manually the domain adaptation for every country. Simply because each country does not experience the same history and can provide more data of one scenario instead of an another. We make the example of Burkina Faso. With the models trained as in the previous section we obtain the following predictions shown in Fig. 6:

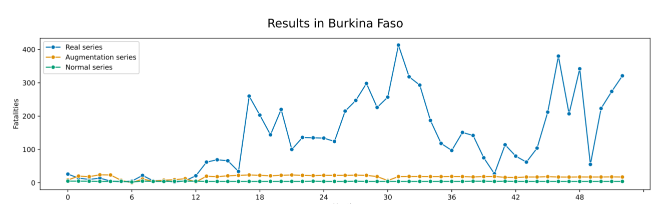


Figure 6: Prediction in Burkina Faso with the split based on political indicators.

We tried to retrain the model, shifting to a more aggressive data augmentation where the tumultuous dataset was sampled starting from data of events exclusively with a number of fatalities larger than 50.

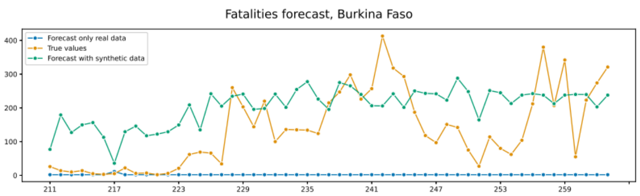


Figure 7: Prediction in Burkina Faso with the split based on fatalities

As we can see in Fig. 7 this split suits the case of Burkina Faso better.

4.3.2 Calibration. For what concerns calibration, we used the validation set to calibrate the models trained in the previous step and then observed the output in the test set. We have noticed that calibration is greatly affected by the choice of the domain split and, in general, augmenting also the calibration set improves the results of calibration, as we show in the plots in figures 8 and 9.

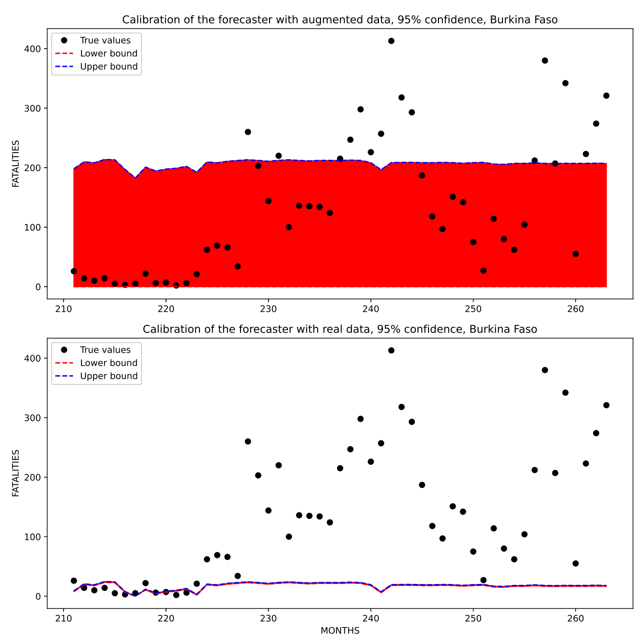


Figure 8: Calibration in Burkina Faso with the split based on political indicators. In the top figure we have causal data augmentation on the calibration set, in the bottom figure not.

We believe that those steps are essentials to ensure good calibration. Well calibrated models can provide strong guarantees on probabilities, driving a better decision making.

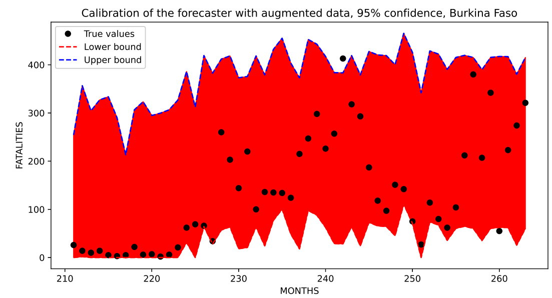


Figure 9: Calibration in Burkina Faso with the split based on fatalities and causal data augmentation on the calibration set.

5 CONCLUSIONS

In this work we have presented the construction of the dataset to predict fatalities in the Sahel region. We have also illustrated Causal Data Augmentation, which provide a flexible, model agnostic methodology to augment data in the case of distribution shifts. Moreover, we have shown how causal data augmentation can be beneficial in the calibration setting. We stress the importance of calibration, as it can provide confidence intervals regarding the likelihood of an event. This is of particular interest in our case as knowing the exact number of fatalities is not crucial, while an interval with strong probabilistic guarantees can provide a good basis to the stakeholders of our project to act and prevent catastrophic events.

5.1 Future Works

Causal Data Augmentation needs refining, as we believe it needs more stability, as it can provide remarkably different results over different runs. Moreover, data generation with CDA is a slow process and we believe it would be beneficial to research ways to speed it up. Another research direction lies in automating the choice of the domain split, as it is not a trivial hyperparameter to tune and the tuning process requires a huge amount of resources and time.

5.2 Social impact and Feedback Loops

In the case a similar model is deployed and employed by organisation to prevent conflicts and hopefully succeed, one must consider that society might adapt to escape such prediction system. One must consider that malicious organisation could act in a way to deceive the predictive model. This, while far-fetched for small groups, is a real possibility for state-level organisations. This motivates us to restrict the complexity of the dataset, following the principle of data minimisation, already present in law, such as in the General Data Protection Regulation [11]. Similarly, one must consider how such a system will affect the economy and society of target countries, especially in long-term predictions. Consider, for instance, a false positive prediction of a war or terrorist attack in a certain country with a 12 months anticipation. With such a long warning, the targeted country could see many investments or events cancelled in fear of some kind of social conflict. This can have a great impact on the freedom and fundamental rights of the people living in the target country, as it hinders the development of such country. This would increase the divide with more developed and peaceful countries, which would likely see the investments and events to be directed to them. In the long term this could cause more instability in the target country, reinforcing the bias. To conclude, we invoke caution in the usage and deployment of predictive models for social conflicts.

REFERENCES

[1] Use of early warning systems in western africa to combat terrorism. 2019.

[2] ACLED. Acled data codebook. 2023. URL https://acleddata.com/knowledge-base/codebook/.

[3] Mohammad Al-Saidi, S.A. Saad, and Nadir Elagib. From scenario to mounting risks: Covid-19's perils for development and supply security in the sahel. *Environment, Development and Sustainability*, 25, 04 2022. doi: 10.1007/s10668-022-02303-9.

[4] Poinsot Audrey and Leite Alessandro. A guide for practical use of admg causal data augmentation. *arXiv preprint arXiv:2304.01237*, 2023.

[5] Various authors. Lingam documentation. https://lingam.readthedocs.io/en/latest/index.html, 2022. Accessed: yyyy-mm-dd.

[6] Various authors. How can i draw a value randomly from a kernel density estimate? https://stats.stackexchange.com/questions/321542/how-can-i-draw-a-value-randomly-from-a-kernel-density-estimate, 2022. Accessed: 2024-01-18.

[7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[8] Mustafa Daraghmeh, Anjali Agarwal, Ricardo Manzano, and Marzia Zaman. Time series forecasting using facebook prophet for cloud resource management. pages 1–6, 2021.

[9] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

[10] Shereen Elsayed, Daniela Thyssens, Ahmed Rashed, Hadi Samer Jomaa, and Lars Schmidt-Thieme. Do we really need deep learning models for time series forecasting? *arXiv preprint arXiv:2101.02118*, 2021.

[11] European Parliament and Council. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), 2016. URL https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[12] Daniella Frias, Carolina Cavour Siqueira Muniz, Pedro Senna Vieira, Dominique Souza Sant'anna, and Augusto da Cunha Reis. Supply chain risk mitigation through sales forecasting in a cosmetics company. *Independent Journal of Management & Production*, 11(5):1606–1623, 2020.

[13] Azul Garza and Max Mergenthaler-Canseco. Timegpt-1. *arXiv preprint arXiv:2310.03589*, 2023.

[14] Håvard Hegre, Angelica Lindqvist-McGowan, Paola Vesco, Remco Jansen, and Malika Rakhmankulova. Forecasting armed conflict in the sahel: Forecasts for november 2021–october 2024. 2022.

[15] Nikolas Roman Herbst, Nikolaus Huber, Samuel Kounev, and Erich Amrehn. Self-adaptive workload classification and forecasting for proactive resource provisioning. pages 187–198, 2013.

[16] Maximilian Ilse, Jakub M Tomczak, and Patrick Forré. Selecting data augmentation for simulating interventions. In *International Conference on Machine Learning*, pages 4555–4562. PMLR, 2021.

[17] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[18] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[19] Steffen L Lauritzen. Causal inference from graphical models. *Monographs on Statistics and Applied Probability*, 87:63–108, 2001.

[20] Šedová B. Schewe J. et al. Michelini, S. Extreme weather impacts do not improve conflict predictions in africa. 2023. doi: https://doi.org/10.1057/s41599-023-01996-1.

[21] Cheolhwan Oh, Seungmin Han, and Jongpil Jeong. Time-series data augmentation based on interpolation. *Procedia Computer Science*, 175:64–71, 2020.

[22] Judea Pearl. Causal inference. *Causality: objectives and assessment*, pages 39–58, 2010.

[23] Akah Pius Odey and Osmond Otora. An assessment of ecowas early warning and crises respond network(ecowarn) and transborder criminality in west africa, 1999-2021. 15:2022, 08 2023.

[24] UNDP (United Nations Development Programme). Human development report 1990. *UNDP (United Nations Development Programme)*, 1990.

[25] Bintu Zahara Sakor. Is demography a threat to peace and security in the sahel? 2022. https://wcaro.unfpa.org/sites/default/files/pub-pdf/en_-_is_demography_a_threat_to_peace_and_security_in_the_sahel.pdf.

[26] Jonas Vestby1 3 Jørgen Carling 1 Halvard Buhaug 1 2 3 Sebastian Schutte 1, 3. Climatic conditions are weak predictors of asylum migration. 2021. doi: https://doi.org/10.1038/s41467-021-22255-4.

[27] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[28] Shohei Shimizu. Lingam: Non-gaussian methods for estimating causal structures. *Behaviormetrika*, 41:65–98, 2014.

[29] Ravid Shwartz-Ziv and Amitai Armon. Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90, 2022.

[30] Bernard W Silverman. Spline smoothing: the equivalent variable kernel method. *The annals of Statistics*, pages 898–916, 1984.

[31] Stockholm International Peace Research Institute (SIPRI). Sipri military expenditure database. URL https://www.sipri.org/databases/milex.

[32] The Armed Conflict Location & Event Data Project. The armed conflict location event data project. 2023. URL https://acleddata.com/data-export-tool/.

[33] Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

[34] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[35] United Nations Development Programme. United nations development programme. 2023. URL https://www.undp.org/.

[36] Ruoyu Wang, Mingyang Yi, Zhitang Chen, and Shengyu Zhu. Out-of-distribution generalization with causal invariant transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 375–385, 2022.

[37] World Bank. World bank databank indexes, 2023. URL https://databank.worldbank.org/.

[38] Mengyue Yang, Furui Liu, Zhitang Chen, Xinwei Shen, Jianye Hao, and Jun Wang. Causalvae: Disentangled representation learning via neural structural causal models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9593–9602, 2021.

A ACLED FORECASTING FATALITIES

A.1 Introduction

In this report we perform a short statistical analysis of the ACLED dataset, for then searching for some potentially good features for a forecasting model.

A.2 Feature Selection

The dataset is characterized by entries of political and social conflicts registered in the African continent between 01-01-1997 and 10-11-2023. Each entry of the dataset represents an event, described by 31 features, of which, only 4 have been considered in our preliminary study, namely:

- *EVENT_DATE*
- *DISORDER_TYPE* which comprehends four types of disorders (demonstrations, political violence, strategic developments and political violence + demonstrations)
- *COUNTRY*
- *FATALITIES* which represents the number of fatalities registered for the specific event, and represents our target variable

These four features ensured a first naive analysis of the fatalities trend due to political and social conflicts in the Sahel region. For every country, we created a smaller view of the dataset, and we also aggregates the fatalities in month in order to lower the number of zeros in the series. After that, we added four more features in the dataset by computing a decay function in the form $e^{-t/\alpha}$ where t is the number of months passed since at least 5 fatalities, and α is the half-life parameter setted to 6 months. The implemented feature set comprehend one target variable corresponding to the number of fatalities, the four decay functions and 12 lag features (a lag feature is simply the target variable shifted by a certain amount of time); for a given time step t, the lag features represent the values of the variable at earlier time steps t-1,t-2,...,t-k, where k is the lag or time delay.

A.3 Model

In this preliminary analysis, we explored a simple XGBoost regressor in order to predict the number of fatalities. Right now, the model underperforms in terms of predicting capabilities as can be seen in Fig. 10, due to the fact that is very simple and the target is full of zeros, that have to be managed with particular techniques (i.e. Zero-Inflated Models, Hurdle Models or Two-Step Approaches involving different strategies for zero predictions and non-zero predictions). Another inconsistency with the predictions is due to the fact, that they are computed based solely on temporal features and fatalities without any geo-political context, which is one of the challenges we have to address. (See Fig. 11).

B UCDP ANALYSIS

B.1 Introduction

In this report we perform a short statistical analysis of the UCDP dataset, delving into feature selection and corrrelation.

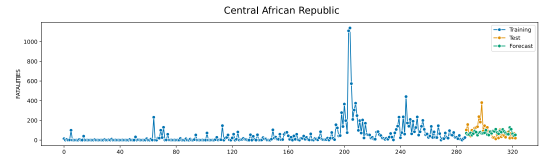


Figure 10: Forecasting for fatalities in Central African Republic between 11-2020 and 11-2023

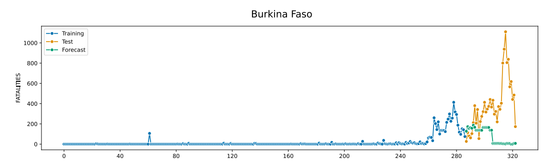


Figure 11: Forecasting for fatalities in Burkina Faso between 11-2020 and 11-2023. Here is clear that a forecasting based solely on fatalities cannot perform well since they depend on external factors and not seasonality or trends.

B.2 Feature Selection

The full description of the dataset is available at https://ucdp.uu.se/downloads/ged/ged231.pdf, we summarise few highlights:

- Each entry of the dataset represents an event, described by 48 features.
- A lot of features are just identification codes.
- Many features have the same information with a different level of granularity: for example country and latitude/longitude.
- Many features are actually pieces of text, like source headline.
- The dataset has few missing data and the features missing data are the textual features and few identifiers.

The first step of our analysis deals with missing data: since the only features with missing values are textual features and some identifiers we can decide to drop them. We opt for dropping them since textual data is cumbersome to manage in our simple analysis and it is better to focus on other kinds of data. Secondly, we choose to exclude geographical in the role of a predictor, since it is a source of biases and likely a confounding factor since it does not directly relate to the event.
After this preprocessing step, we are left with the following features:

- *number of sources*: the number of sources reporting the event.
- *where prec*: the precision of the information of where the event happened as an integer number (the lower, the better precision).

- *priogrid gid*: the identifier of the priogrid, a 55x55 square kilometers region.

- *country*: country where the event happened.

- *event clarity*: it can be 1 or 2, where 1 means direct reporting and 2 means indirect reporting.

- *date prec*: the precision of the information of when the event happened as an integer number (the lower, the better precision).

- *date start*: earliest possible starting date of the event.

- *deaths a*: number of deaths in side a of the conflict.

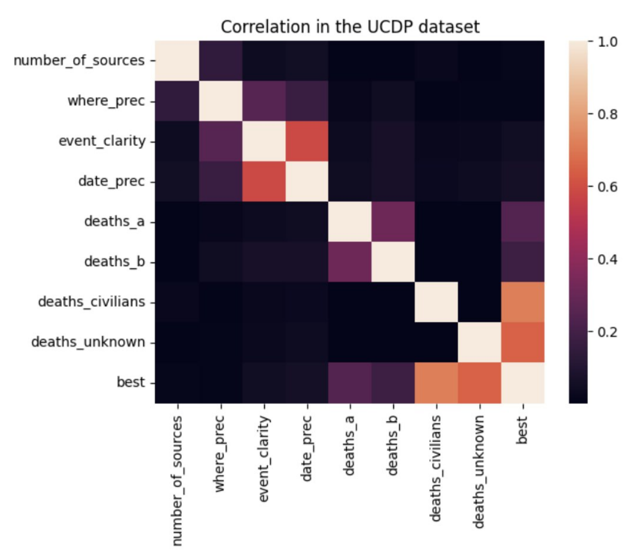
- *deaths b*: number of deaths in side b of the conflict (zero if one-sided conflict).

- *deaths civilian*: number of civilian deaths.

- *deaths unknown*: deaths of persons of unknown status.

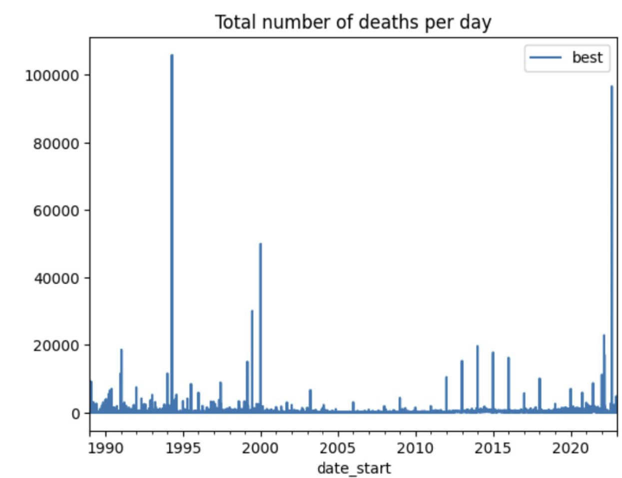
- *best*: best estimate of deaths, it is the sum of deaths a, deaths b, deaths unknwon, deaths civilian.

When building the model, it is better to leave out geographical information to avoid bias and use it only for aggregation and employ it for region-specific models to get more accurate results. We also report the information of the absolute correlation, as we can see deaths do not correlate with non-death features, while we have a high correlation between total deaths and civilian deaths, from this we could infer that a large number of victims in events are civilians.

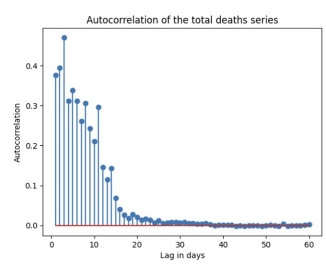


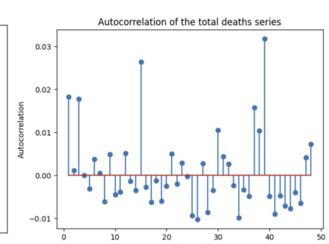
B.3 The time dimension

For a predictive algorithm, the time dimension is of paramount importance. We focus our attention on the total deaths time series. As we can see, this time series has a lot of spikes and it is hard to distinguish patterns. It is also worth considering that biases in reporting may concentrate many casualties in a single day, making the analysis more difficult. For example: two events that could not happen in the same day may have the same possible earliest day and in the reporting they will result in the same day. Nevertheless,



we can consider how deaths correlate time-wise. We perform this analysis by grouping per month and per day. We can see that deaths from month to another do not have a high correlation, whereas we can see how well they correlate in days. This makes sense, since conflicts usually have surges of violence which last a few days.





B.4 Conclusion

We have performed a simple analysis on UCDP, we have seen how well they correlate in the span of days. This has consequences, in the sense that it is harder to justify some kinds of assumptions usually employed in machine learning models. For instance, we cannot assume independent, identically, distributed (iid) samples along time. Similarly, it is harder to fit a Markov Chain model, since we could have long-term dependencies. Those assumptions are at the basis of many modern machine learning and deep learning techniques. Some models, instead require the iid assumption over a finite-length sequence, however it is hard to detect seasonality in UCDP. In practice, these kinds of models are used anyway, but in any case we require a certain degree of robustness. A similar issue holds for calibration strategies, which require exchangeable data points. We believe that those issues are crucial, since the violations of those assumptions make it harder to establish a theoretical guarantee in our predictions.

C CAUSAL DISCOVERY

C.1 Introduction

In this work we explore causal inference and try to apply it to early warning systems.

C.2 Theoretical background

In this subsection we formalise the task and introduce the main definitions

C.2.1 Structural Causal Models. Structural Causal Models (SCMs) are a tool to model causal relationships. The intuition is that a factor V follows from a causal model $pa(V)$ plus an uncertainty factor η. More formally:

$$V = f(pa(V), \eta) \tag{3}$$

Where f is a function, potentially non-linear, which determines the type of causal relationship between a factor and its causal parent. In our domain, causal parents can be previous values of the time series. For example, we can take fatalities as the factor to examine, then we can suspect that there is a causal relationship between the trend of violent events and the magnitude of the next violent event. However, we are not limited to this case, causal parents can be multiple and of different type. As an example, intuitively, we can link a military coup with an increase of violent events due to suppression of resistance.

Causal relationships can be represented with Directed Acyclic Graphs (DAGs), called causal graphs. We need acyclicity due to the principle of cause and effect.

C.2.2 Interventions. With interventions we force some variables to assume some values and discard their causal nature:

$$V = f(pa(V), \eta) \rightarrow V = v \tag{4}$$

An intervention has the effect of removing the arrows in the causal graph that point to V. As an example, we could want to study how government forms influence the occurrence of violent events and we might want to set the government form to be a military dictatorship and study how it affects violent events.

More in general, an intervention on a set of factors X is denotes by $do(X = x)$. Crucially, $P(\cdot|do(X = x)) \neq P(\cdot|X = x)$. This is because interventions force an initial value, but, differently from conditioning, they allow it to change in following time instants according to causal relationships. As an example, suppose we fix a government type to be a republic and study the effect of this. With intervention, we allow the government type to change again if causality allows it. Conditioning, instead, does not allow that.

Interventions model a what if, applying only a divergence point and allowing story to make its course.

C.2.3 Counterfactuality. Counterfactual experiments consist in two steps:

(1) Perform an intervention $do(X = x)$ and observe the target variable Y to be some y
(2) Perform an intervention $do(X = x' \neq x)$ and study the probability of observing $Y = y' \neq y$

C.3 Learning with causal graphs

Causal graphs can be either constructed through domain knowledge or learned from data. The second approach is called *causal inference.* There is plenty of research in learning causal graphs, therefore we shift our attention on how to use the prior knowledge of causal graphs in predictive models and why we should use it.

C.3.1 Motivations. Time series data often undergo domain shifts as we already mentioned in previous subsections. Just think how a war can change the security landscape of a country, impacting homicides and other violence.

Machine learning models trained on a source domain S are not expected to work properly on a different target domain T, therefore we must make adjustment to make the model more robust.

Causal graphs retain a reasoning structure which maintain a different information than the correlations in a dataset, which might change depending on a domain. With this information we can help the machine learning model understand domain shifts and adjust the predictions.

C.3.2 Causal graphs for data augmentation. Augmenting with limited knowledge We can use causal graphs to augment the samples in our dataset by using only the fact that some features are connected to others in some way, without specifying the type of relation. Consider our dataset X, with n samples and d features, we denote by x_i the i-th sample of the dataset and by x_i^j the j-th feature of the i-th sample of the dataset. Then we generate new data points by selecting all possible combinations of the values of the features of the data points. For example we can generate a new sample by taking the first feature from the third sample, the second feature from the first sample and so on. However not all data points generated in that way are reasonable, therefore we score them according to the likelihood that they exist and remove the one that do not score beyond a certain threshold. This estimate can be done through kernel estimation, a non-parametric statistical technique.

Augmenting with relational knowledge If the type of relation between variables in our data set can be estimated, we can leverage interventions to generate alternative plausible scenarios, which can give us probable data points. In fact, this approach generates 'what ifs', which can be thought of as alternative timelines. Moreover, we believe that this approach can generate a higher domain generalisation because it can generate new values not present in the data set. For example, think of a country which has never seen war, then the training data set of that country will not contain events with a high number of fatalities. However, if the causal graph can model the information of warfare, then it will be able to generate alternative data points with a high number of fatalities. This is not possible to do without relational knowledge.

C.3.3 Causal graphs for constrain enforcement. We can use causal graphs to model constraints, for example we know for a fact that the number of deaths in a violent event cannot be lower than zero. In principle machine learning models do not retain this information and it must be provided with a post-processing step. However, causal graphs can help us model such constraints and potentially use them in forecasting.

C.4 Challenges

One challenge of the proposed approach is to get a causal graph which as informative as possible. This can be done by causal discovery methods or by domain knowledge.
Another challenge is to effectively use the topology of the causal graph with boosted trees, since so far we have not used it directly but just to compute new data points or enforce some kind of constraints.

C.5 Conclusion

In future work we aim at increasing the amount of information we can extract from causal graphs, possibly using them at inference time.

D CAUSAL GENERALIZATION

D.1 Introduction

In this log we highlight a research possibility that represents an alternative with respect to Causal Data Augmentation. This leverage training a model for learning distribution shifts, instead of exploiting conditional distributions.

Causal graphs allow us to augment our data set. However, methodologies that we explored so far only allow us to use existing values in our data set. This limits the possibilities of generalising to unseen distributions.
Other approaches exploit prior knowledge to design causal invariant transformations (CITs). A CIT is a function which, when applied on a sample x, does not change its target y. While in fields like computer vision CITs are straightforward to devise, like rotation or Gaussian blurring, in the time series domain it is not trivial.

D.2 Problem specification in the PMHA project

In the PMHA setting we have time series data of multiple countries. We assume that this data is sampled from a unique data distribution X with probability density f_X, then we can subdivide the data by country, obtaining conditional distributions X_c, or by some other kind of status X_s. In our case, training a global model on X can have the effect of being too general and losing the details of the country. However training only on country data gives us not robust models. Data distributions are characterised by: country and stability status. While the country always stays the same, its stability can change. Hence we observe the following shift:

$$X_{c,s_1} \rightarrow X_{c,s_2}$$

Ideally, for a country we want to train on data that covers both cases but it is not always possible because country might have never been at war for example, or at least just for a very short time. However, if we considering all countries, we have a large amount of data regarding all of the stability statuses. Therefore, we know:

- **The conditional distribution over stability status** $P(X = x|S = s)$, presumably for all statuses
- **The conditional distribution over stability status and country** $P(X = x|S = s, C = c)$ but just for some s, c pairs

We want $P(X = x|C = c)$. If we assume that countries and stability statuses are independent, we can write:

$$P(X = x|C = c) = \sum_s P(X = x|C = c, S = s)P(S = s)$$

But we do not know $P(X = x|C = c, S = s)$ for all s. However, if we know $P(X = x|S = s)$ for all s, it is reasonable to assume that we can learn some function $g_{i,j}$ such that:

$$P(X = x|S = s_i) = P(X = g_{i,j}(x)|S = s_j) \ \forall \ i, j, x \sim X$$

But does this guarantee $P(X = x|S = s_i, C = c) = P(X = g_{i,j}(x)|S = s_j, C = c)$? Unfortunately not. However, in our setting we can assume that the error is probably limited:

$$|P(X = x|S = s_i, C = c) - P(X = g_{i,j}(x)|S = s_j, C = c)| < \epsilon$$

This is because countries behave similarly in a war, for example we usually do not see a decrease in mortality during wars. How can we learn g? We can minimise:

$$\underset{g \in \mathcal{H}}{\operatorname{argmin}} |P(X = x|S = s_i) - P(X = g_{i,j}(x)|S = s_j)|$$

To estimate both probabilities we can use kernel methods as we did previously. This method generalises CITs and allows us to perform a stronger domain shift.

D.3 Exploiting graphical causal knowledge

If we have a better graphical knowledge, we can split our covariates in two groups:

- D: variables dependent from S
- I: variables independent from S

This additional knowledge can allow us to employ the model only for D and generating I in some other way:

$$\underset{g \in \mathcal{H}}{\operatorname{argmin}} |P(D = d|S = s_i) - P(D = g_{i,j}(d)|S = s_j)|$$

We have to note that, even though this helps reducing the dimensionality of the task, we still have a multi-output regression. Which can be challenging in many cases. One solution relies on discarding interactions between output variables, essentially assuming they are independent. However this is a strong assumption and its expected violation can lead to poor results.

D.3.1 Interventions for data augmentation. In this case graphical causal knowledge comes into play again. We first consider the set of independent variables I and we either sample from there or we directly take values from the training data. Then we use the data that we obtained previously, together with the shift $s_i \rightarrow s_j$ and we perform an intervention on the causal graph. Finally, we capture the changes using the dynamics of the graph and use them as augmented data.
The dynamic of the graph can be simulated using expectations over a conditional distribution obtained with kernel methods over the distribution $P(D = d|S = s_j, I = i)$.

E FORECAST OUTSIDE OF SAHEL

E.1 Introduction

In this log we apply the Causal Data Augmentation techniques outside of Sahel, to see how they perform.

E.2 Discussion

We implemented the techniques using a split over a political indicator. We apply our method to Ukraine, Ecuador and Tunisia. Notice that our data stops before the recent tensions in Ecuador. We notice that causal data augmentation struggles to improve the forecasting in these countries, we believe that this can be motivated by the following points:

- The political indicator could not be a good split for these countries.
- Tunisia and Ecuador do not perform a distribution shift, hence we pay for a generalization that is not needed.
- Ukraine has a distribution shift, but the political indicator is not the best split for it, as it enters a state of war and the split of the political indicator is not a good split, as it does not characterise the violence of a full-scale war. It is a similar situation to the one in Burkina Faso, where the political indicator is not a good split for that situation.

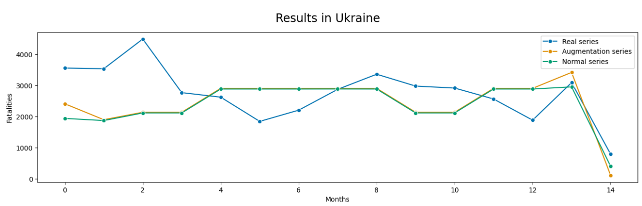


Figure 12: Results for Ukraine.

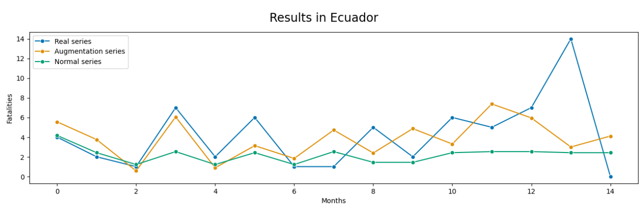


Figure 13: Results for Ecuador.

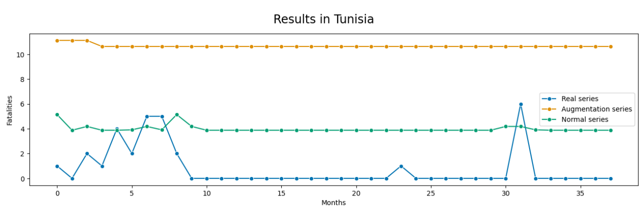


Figure 14: Results for Tunisia.

E.3 Conclusion

With this short log we highlight the challenges and the future work that could improve CDA. We also believe that this work can be useful for underlying the importance of the split in the CDA method and the difficulties of forecasters in general.