Summary for the 2023/24 VIEWS Prediction Competition*

Tobias Bodentien[1] and Lotta Rüter[2]

[1] *Institute of Statistics (STAT), Karlsruhe Institute of Technology (KIT), tobias.bodentien@student.kit.edu*
[2] *Institute of Statistics (STAT), Karlsruhe Institute of Technology (KIT), lotta.rueter@kit.edu*

June 27, 2024

Abstract

We choose a data-driven approach for our contribution to the 2023/24 VIEWS prediction competition. We focus on the country-month (cm) level and base our analysis on the data provided by the VIEWS team. Since it was communicated as the main metric in advance, our primary objective is to achieve optimal predictive performance with respect to the Continuous Ranked Probability Score (CRPS). We compare three modeling approaches that differ in their levels of complexity: a negative binomial distribution (NB), a hurdle model and feed-forward neural networks (NNs). Our model determination is based on test data that was available at the start of the challenge, i.e. the years 2018 to 2021. Despite being the simplest of the three models, we find that the NB outperforms the other two more involved approaches in terms of the average CRPS in these years, while also being competitive or superior in the challenge's secondary metrics. We therefore submit forecasts generated by a NB. More specifically, we use its $0.1, ..., 99.9\%$-quantiles as our predictive samples. The resulting model is simple, straightforward, transparent and easy to interpret. It naturally models the conflict trap characteristic, yet by construction it is unable to predict the outbreak of conflicts or identify trends that have not occurred in the past. Regarding the additional test data for 2022 and 2023 that was released closer to the submission deadline, we find the NNs to be superior to our simpler approaches, which shows that they provide a promising path for potential future extensions of our work.

VIEWS PREDICTING CONFLICT & HUMANITARIAN IMPACTS

Federal Foreign Office

PREVIEW Prediction Visualisation Early Warning

*This paper documents a contribution to the VIEWS Prediction Challenge 2023/2024 (Hegre et al. 2023). Financial support for the Prediction Challenge was provided by the German Ministry for Foreign Affairs. For more information on the Prediction Challenge please see Hegre et al. (Forthcoming) and viewsforecasting.org

1 Introduction

Compared to other approaches that add uncertainty estimates to a point estimate to move from a point prediction to a probabilistic forecast, we model the predictive distribution of *cm*-fatalities directly. For model selection purposes, we compare three approaches that differ in their levels of complexity and their flexibility to incorporate specific characteristics inherent in the data, see Table 1 for an overview.

	NB	Hurdle Model	Neural Networks
Overdispersion	✓	✓	✓
Zero-Inflation		✓	✓
Spatio-Temporal Dependencies			✓
Complexity	low	middle	high

Table 1: Characteristics of our three modeling approaches and their complexity.

First, we utilize a negative binomial distribution (NB) to account for the overdispersion inherent in the data. Its parameters are estimated via empirical moments of the country's past w fatalities. By construction, this approach is unable to predict conflict onset. Second, we employ a hurdle model that additionally accounts for zero-inflation by modeling the distribution of zeros separately using a Bernoulli variable. Positive numbers of fatalities are modeled via a truncated negative binomial distribution. Again, the respective model parameters are estimated based on past fatalities. Third, we flexibly incorporate additional feature variables provided by the VIEWS team using feed-forward neural networks to further model spatio-temporal dependencies. In all three cases, we tune the hyperparameters in such a way that the average CRPS is minimized. The same holds for the ensuing model selection.

This summary is structured as follows. We first review the characteristics of the data and the resulting challenges for the prediction task in Section 2. We then present our three modeling approaches in Section 3, followed by a description of our results in Section 4. Section 5 concludes.

2 Data Characteristics

We base our predictions solely on the data provided by the VIEWS team. At the *cm*-level, these data comprise monthly fatalities as well as observations of 123 additional feature variables for 191 countries, 93 of which have zero recorded fatalities as of October 2022. The prediction target, that is, the number of fatalities due to state-based conflict, exhibits certain characteristics that require appropriate modeling strategies. First, the data are count data and hence integer-valued. Second, they are overdispersed meaning that the variance in the data is often higher than expected by a simple model, for instance, a Poisson distribution. Third, due to many non-conflict countries and peaceful months, the data are (fortunately) zero-inflated. The additional feature variables, comprising lagged, spatial conflict data, aspects of democracy and development indicators, amongst others, are divided into three main categories *conflicthistory*, *vdem* and *wdi*, see Section 4.1.1 in VIEWS team (2022).

For the visualization of the results, we divide all 98 countries with at least one month of non-zero reported fatalities into three conflict levels depending on the average number of fatalities $\overline{Y}_i$ observed until October 2022:

1. **Low conflict level** ($0 < \overline{Y}_i \leq 5$), 50 countries
2. **Moderate conflict level** ($5 < \overline{Y}_i \leq 100$), 38 countries
3. **High conflict level** ($\overline{Y}_i > 100$), 10 countries

Figure 1 yields an illustration of the different categories.

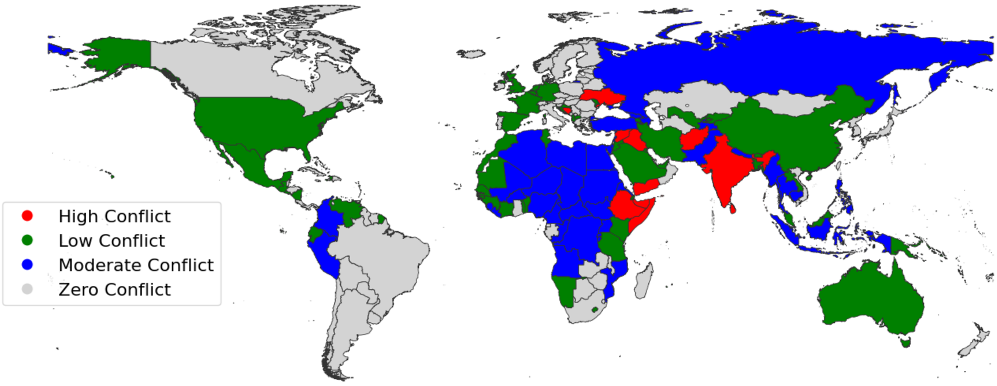


Figure 1: Categorization of countries into zero, low, moderate, and high levels of conflict.

3 Modeling Approaches

The underlying concepts of our three modeling approaches are presented below while technical details on the modeling and estimation process can be found in Appendix A.

3.1 Negative Binomial Distribution

The negative binomial distribution has been deployed in various areas of research to model overdispersed count data, for example, in sociology, epidemiology and ecology (Moghimbeigi et al. 2008; Ver Hoef and Boveng 2007). We estimate its parameters from the past w observations as described in Appendix A.1. Since w is unknown, we consider $w \in \{2, ..., 24\}$ and determine its optimal value as the minimizer of the average CRPS across all countries i and months t contained in the training data of Task 2. The CRPS for observation $y_{i,t+s}$ is computed based on the 0.1%, 0.2%,...,99.9%-quantiles of the respective negative binomial distribution for window length w. To ensure that we base our analysis of different values of w on the same amount of data per subtask (2018 to 2023), we limit our analysis to the five most recent years available in the respective training dataset. As shown in Table 2, we consider four different variants for determining the optimal w.

Variant	Minimize the average CRPS over one w	Optimal w
1	For all countries	w_1^*
2	For each country i	w_{2i}^* (for $i = 1, ..., 191$)
3	For each period s	w_{3s}^* (for $s = 3, ..., 14$)
4	For each country i and period s	w_{4is}^* (for $i = 1, ..., 191$ and $s = 3, ..., 14$)

Table 2: The four different approaches we use to obtain the optimal w.

3.2 Hurdle Model

As an extension to the NB approach, we estimate a hurdle model which was first described by Cragg (1971) to additionally account for the excess zeros in the data. In contrast to the NB, the hurdle model contains an additional Bernoulli variable that separately models the chance of non-zero fatalities per country. Positive numbers of fatalities are modeled by a truncated negative binomial distribution (TNB). Again, technical details are given in the Appendix A.2 and the optimal window length w^* is determined as described in Table 2.

3.3 Neural Networks

To investigate the benefit of incorporating additional feature data using a flexible, data-driven model, we compare the above two methods to feed-forward neural networks.

We model the predictive distributions of fatalities by training separate individual NNs for each country $i = 1, ..., 191$ and lead time $s = 3, ..., 14$. As in the other two models, we use the number of conflict deaths in the last w months as inputs and additionally include a feature set f_{set}. We set the number of output neurons to $n_{\text{output}} = 200$, each representing a draw from the predictive distribution. To ensure that the predicted values are non-negative and integer-valued, the outputs are ReLU-transformed and rounded to integers. For CRPS optimality, we employ the energy form of the metric as the loss function (Gneiting and Raftery 2007, p. 367). The hyperparameters $(w, h, N_h, l, b, d, e, f_{\text{set}})$, all defined in the following, are determined via a random search of 20 trials per country, see Bergstra and Bengio (2012), with the maximum number of previous months used for prediction denoted by w, the number of hidden layers h, the total number of neurons in the hidden layers N_h, learning rate l, batch size b, dropout rate d, epoch size e and feature subset f_{set}. They are tuned only for countries with non-zero fatalities. For countries with no reported fatalities, we issue a sample of only zeros. In all other cases, one neural network per country and lead time $s = 3, ..., 14$ is trained with the selected hyperparameters on the respective training data of Task 2.

The training data for each subtask of Task 2 are split into 70% training and 30% validation data, as commonly done in practice (Joseph 2022, p. 531), where the most recent data are used for validation. For more details on the network characteristics and tuning procedure, please refer to Appendix A.3.

4 Results

4.1 Optimal window length for the NB and the Hurdle Model

The distinction between the negative binomial distribution and the hurdle model as well as the four variants of determining w^*, see Table 2, results in eight different models. Averaging over all countries and all six test windows 2018–2023 yields the CRPS values presented in Table 3. With regard to w, it turns out that for countries with non-zero conflicts, larger values tend to be optimal in all variants of the baseline. For countries with no reported conflicts, the choice of w does not affect the results; to maintain consistency, we choose the smallest w, namely $w = 2$.

Results 2018–2023	**NB** Average CRPS	$(\bar{w}^*,\ sd_{w^*})$	**Hurdle Model** Average CRPS	$(\bar{w}^*,\ sd_{w^*})$
Variant 1 with w_1^*	56.283	(16.5, 8.4)	**56.662**	(16.5, 8.4)
Variant 2 with w_{2i}^*	77.112	(5.93, 7.83)	77.470	(5.9, 7.67)
Variant 3 with w_{3s}^*	**56.110**	(18.62, 7.92)	58.650	(17.85, 8.55)
Variant 4 with w_{4is}^*	69.246	(5.52, 7.36)	74.685	(5.32, 6.98)

Table 3: Average CRPS of the eight different baseline variants. The CRPS is averaged over the six test windows 2018–2023. In addition, the empirical mean $\bar{w}^*$ and standard deviation sd_{w^*} of the optimal w^* of the 191 countries and six test windows are shown.

We find that the more elaborate hurdle model is not able to outperform the negative binomial model in any of the four variants regarding the average CRPS. In case of the NB model, optimizing w for each lead time $s = 3, ..., 14$ yields the best results, while an additional discrimination for each country does not lead to further improvements. The overall optimal baseline for Task 2 is NB Variant 3 while the optimal hurdle model is given by Hurdle Variant 1.

4.2 Model Performance

The results of our models are presented in Table 4 along with those of two VIEWS baselines: the Conflictology benchmark (`bm_conflictology_country12`) and the Last Poisson benchmark (`bm_last_historical`). See Hegre et al. (2023) for a description of these models. An overview of the optimal model per country with regard to the average CRPS per year is given in Figure B.1 in the Appendix.

We find that each of our three models is able to beat the VIEWS Last Poisson benchmark on average over all years and for each error metric. However, the Conflictology benchmark is the best model on average in terms of CRPS (49.36) and MIS (873.53). The NB Variant 3 is only slightly worse regarding the overall average CRPS (56.11) and produces the best IGN (0.61), both individually and aggregated. In terms of MIS, it produces the lowest scores in two out of six years. The more advanced Hurdle Variant 1 is not optimal with regard to the CRPS and IGN in any of the six years.

The most complex approach, the NNs, that are the minimizer of the overall average CRPS of our three models (52.72), fail to outperform the VIEWS Conflictology benchmark (49.36). If we abstract from the CRPS and MIS for 2022 and 2023, we find that all other metrics yield higher values for the NNs than for the NB Variant 3 and the Hurdle Variant

Model	Metric	2018	2019	2020	2021	2022	2023	2018–23
VIEWS Conflictology	CRPS	14.483	**9.146**	21.339	76.850	123.995	**50.357**	**49.362**
	IGN	0.640	0.610	0.567	0.686	0.695	0.682	0.647
	MIS	186.554	**89.058**	344.964	1435.555	**2142.128**	1042.916	**873.529**
VIEWS Last Poisson	CRPS	20.174	9.480	23.698	85.606	131.017	678.960	158.157
	IGN	1.198	1.046	1.110	1.228	1.124	1.125	1.139
	MIS	380.623	172.686	455.806	1690.711	2599.278	13523.463	3137.095
NB Variant 3	CRPS	**14.084**	11.129	**20.640**	**76.577**	125.616	88.614	56.110
	IGN	**0.598**	**0.558**	**0.545**	**0.634**	**0.679**	**0.670**	**0.614**
	MIS	**145.102**	113.754	**317.478**	1405.207	2177.885	1234.818	899.040
Hurdle Variant 1	CRPS	14.360	11.126	20.682	77.012	128.842	87.951	56.662
	IGN	0.752	0.654	0.651	0.728	0.805	0.783	0.723
	MIS	146.887	112.319	318.610	**1398.950**	2309.887	1245.483	922.023
Neural Network	CRPS	16.529	16.236	25.293	81.838	**121.242**	55.157	52.716
	IGN	1.072	1.061	1.072	1.147	1.157	1.178	1.115
	MIS	227.523	229.862	406.504	1548.841	2370.910	**1038.628**	970.378

Table 4: Average error metrics for our three models and two VIEWS benchmarks across the forecast periods of Task 2 and all countries. The right column yields the average across all years.

1, both per year and across the different forecasting horizons. This overall behaviour also holds true for monthly CRPS values for 2018 to 2021, see Figure 2. Even when further distinguishing between different levels of conflict as shown in Figure B.2, we find no clear sign of the NNs being superior in the majority of the considered situations.

The differences in CRPS values are primarily due to a general difference in CRPS levels across all months, rather than differing behaviour during specific time periods or at varying conflict levels.

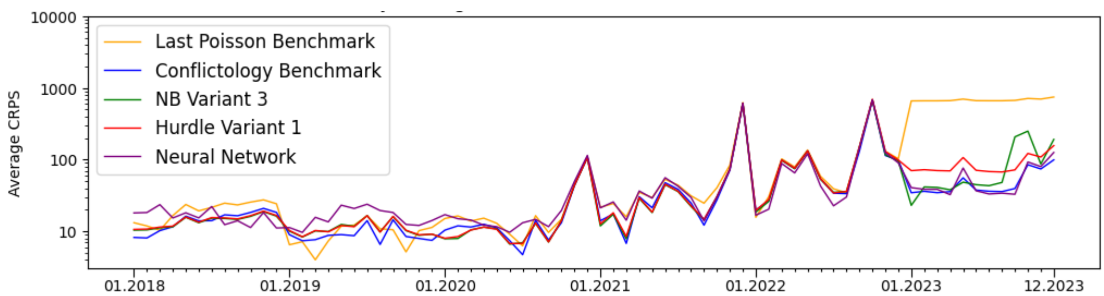


Figure 2: Average CRPS values per model and month in Task 2.

Since we based our model selection on data available in the beginning of the prediction competition, i.e. the years 2018 to 2021, we submit probabilistic forecasts generated by NB Variant 3 for the true future in 2024/25. Additionally, NB Variant 3 is favored for its straightforward interpretability and better performance with respect to the IGN. In Hegre et al. (2023) the NB Variant 3 is marked as `bodentien_rueter_negbin`.

5 Conclusion

We model the predictive distribution of monthly fatalities on a country-level by a negative binomial distribution, a hurdle model that utilizes a truncated negative binomial distribution in combination with a Bernoulli distributed variable to separately account for excess zeros, and feed-forward neural networks. The first two models are estimated solely based on past observations of fatalities of the country the predictions are issued for. The neural networks incorporate additional feature data provided by the VIEWS organizers, including spatio- and temporal lags of fatalities as well as economic and development indicators, amongst others. In total, we thereby successively account for the overdispersion, zero-inflation and spatio-temporal dependencies within the conflict data. We optimize for the CRPS and find that none of our models is able to outperform the VIEWS Conflictology benchmark regarding the average CRPS across all test years. Selecting our optimal model based on the data available at the beginning of the competition, i.e. the years 2018 to 2021, we compute predictions for the true future using the NB model which offers additional advantages. It is straightforward, simple, and transparent, and it outperforms the NN in terms of the secondary metrics IGN and MIS. Still, the NNs perform best in terms of the average CRPS for 2018 to 2023. They thereby represent a promising approach that could be further improved through various enhancements, such as integrating additional data from alternative sources.

References

Bergstra, James and Y. Bengio (Mar. 2012). "Random Search for Hyper-Parameter Optimization". In: *The Journal of Machine Learning Research* 13, pp. 281–305.

Bliss, Chester Ittner and Ronald A Fisher (1953). "Fitting the negative binomial distribution to biological data". In: *Biometrics* 9.2, pp. 176–200.

Cragg, John G (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods". In: *Econometrica* 39.5, pp. 829–844.

Gneiting, Tilmann and Adrian Raftery (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association* 102, pp. 359–378. DOI: 10.1198/016214506000001437.

Hegre, Håvard et al. (2023). "The 2023/24 VIEWS prediction competition". In: *Journal of Peace Research*. Forthcoming.

Joseph, V. Roshan (2022). "Optimal ratio for data splitting". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 15.4, pp. 531–538. DOI: https://doi.org/10.1002/sam.11583.

Kingma, Diederik P. and Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs.LG].

Lindén, Andreas and Samu Mäntyniemi (2011). "Using the Negative Binomial Distribution to Model Overdispersion in Ecological Count Data". In: *Ecology* 92.7, pp. 1414–1421.

Moghimbeigi, Abbas et al. (2008). "Multilevel Zero-Inflated Negative Binomial Regression Modeling for Over-Dispersed Count Data with Extra Zeros". In: *Journal of Applied Statistics* 35.10, pp. 1193–1202.

Porter, Michael D. and Gentry White (2012). "Self-Exciting Hurdle Models for Terrorist Activity". In: *The Annals of Applied Statistics* 6.1, pp. 106–124. ISSN: 19326157, 19417330.

The VIEWS team (2022). *Forecasting Fatalities*. URL: https://www.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf.

Ver Hoef, Jay M and Peter L Boveng (2007). "Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?" In: *Ecology* 88.11, pp. 2766–2772.

Appendix A Modeling Details

A.1 Negative Binomial Distribution

We use the negative binomial distribution to model the predictive distribution of the s-step ahead number of fatalities $Y_{i,t+s}$ of country i issued at time t. The NB is characterized by two parameters, r and p. Following the parameterization from Lindén and Mäntyniemi (2011), $Y \sim NB(r, p)$ implies that

$$P_{\text{NB}}(Y = y \,|\, r, p) = \binom{y + r - 1}{y} p^r (1 - p)^y,$$

where the parameters can be expressed in terms of the expected value μ and variance σ^2 of Y, i.e.

$$r = \frac{\mu^2}{\sigma^2 - \mu} \tag{1}$$

and

$$p = \frac{\mu}{\sigma^2}. \tag{2}$$

As shown by Bliss and Fisher (1953), we can estimate μ and σ^2 via the mean and the empirical variance of Y and use (1) and (2) to obtain estimates for r and p. In our case, we estimate $\mu_{i,t+s}$ and $\sigma^2_{i,t+s}$ from the previous w observations $\{y_{i,t-w+1}, ..., y_{i,t}\}$ of $Y_{i,t}$ via

$$\hat{\mu}^w_{i,t} = \frac{1}{w} \sum_{l=t-w+1}^{t} y_{i,l}$$

and

$$\hat{\sigma}^{2,w}_{i,t} = \frac{1}{w} \sum_{l=t-w+1}^{t} (y_{i,l} - \hat{\mu}^w_{i,t})^2.$$

Plugging these into (1) and (2) leaves us with a fully specified probability distribution for $Y_{i,t+s}$. We determine w in a data-driven way, see Section 3.1.

A.2 Hurdle Model

In contrast to the NB approach, the hurdle model contains an additional Bernoulli variable $Z_{i,t+s}$ whose parameter $\pi_{i,t+s}$ denotes the probability of a non-zero number of fatalities for country i at time $t+s$. In that way, zero fatalities are accounted for separately and positive numbers of fatalities are modeled by a truncated negative binomial distribution. Omitting the indices $i, t + s$, the probability mass function of the hurdle model is given by

$$P_{\text{H}}(Y = y) = \begin{cases} P_{\text{TNB}}(Y = y \,|\, r, p) \cdot \pi, & y > 0, \\ 1 - \pi, & y = 0, \end{cases} \tag{3}$$

see Porter and White (2012, p. 111), where, in our case, P_{TNB} is the probability mass function of the truncated negative binomial distribution.The cumulative distribution function of the TNB is then

$$P_{\text{TNB}}(Y \leq y \,|\, r, p) = \frac{P_{\text{NB}}(Y \leq y \,|\, r, p) - P_{\text{NB}}(Y = 0 \,|\, r, p)}{1 - P_{\text{NB}}(Y = 0 \,|\, r, p)}.$$

We estimate $\pi_{i,t+s}$ via $\hat{\pi}_{i,t+s} = \frac{1}{w}\sum_{l=t-w+1}^{t} \mathbf{1}\{y_{i,l} > 0\}$, the relative occurrence of positive fatality counts in a window of w past observations. The parameters of the TNB are calculated as described in the previous Section A.1, where only positive values in the last w numbers of fatalities are used for estimation of $\mu_{i,t+s}$ and $\sigma^2_{i,t+s}$.

A.3 Neural Networks

The hidden layers of the NNs each utilize the ReLu activation function $f(x) = \max(0, x)$. We choose the number of neurons per hidden layer n_j to be the same for all h hidden layers $j = 1, ..., h$ and the total number of neurons in the hidden layers $N_h = \sum_{j=1}^{h} n_j$ to lie in the interval $[\min(n_{\text{input}}, n_{\text{output}}), \max(n_{\text{input}}, n_{\text{output}})]$, where n_{input} (n_{output}) denotes the number of input (output) neurons.

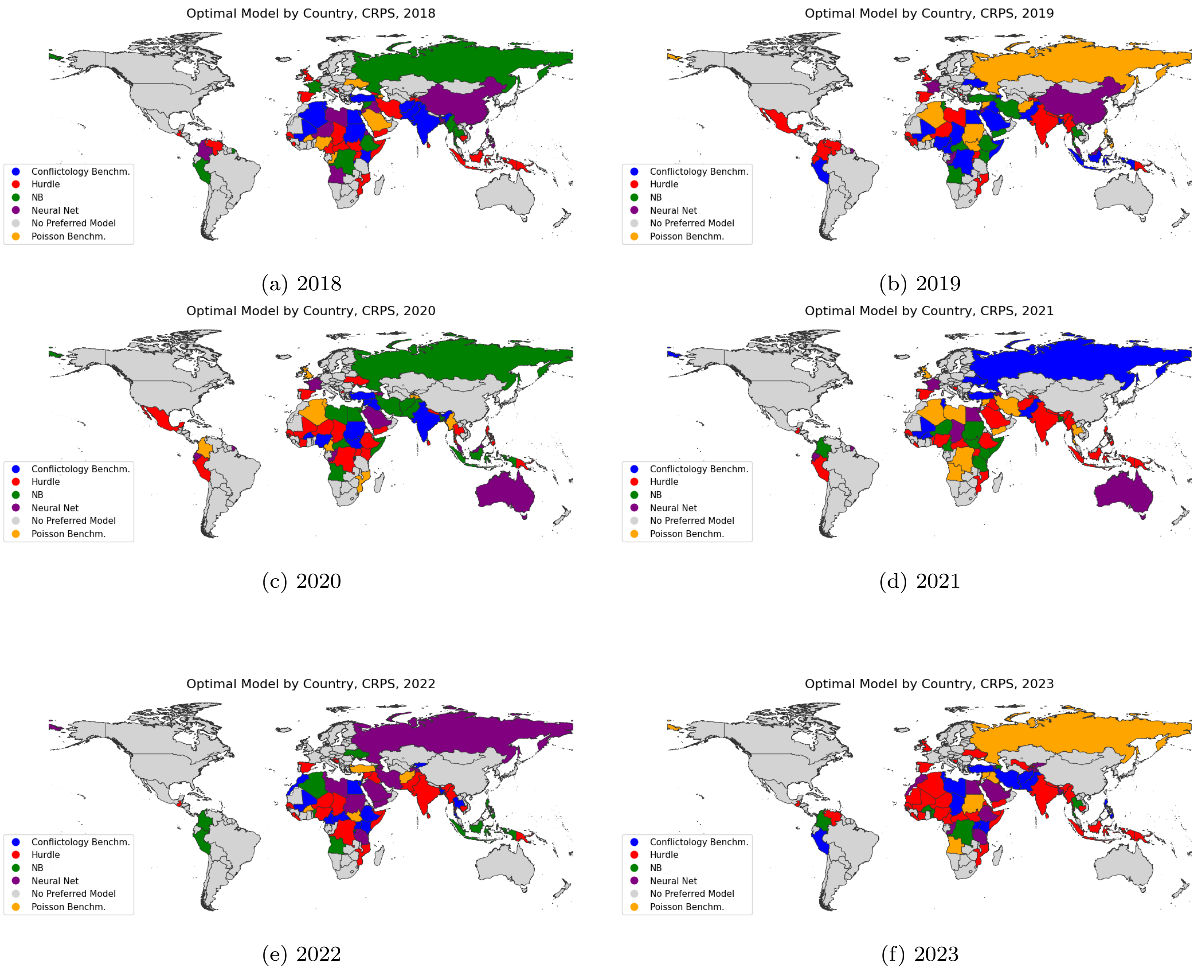
As mentioned in Section 3.3, hyperparameter optimization is performed through random search. In each iteration, the parameter combination to be examined is randomly sampled from the distributions shown in Table A.1. For each country, the chosen set of hyperparameters is that with the minimum CRPS on the validation dataset.

Parameter	**Parameter Space**
w	$U_d(1, 12)$
h, b	$U_d(1, 6)$
N_h	$U_d([\min(n_{\text{input}}, n_{\text{output}}), \max(n_{\text{input}}, n_{\text{output}})])$
l	$U(0.001, 0.15)$
d	$U(0.1, 0.5) \times Ber(0.5)$
e	$U_d(3, 40)$
f_{set}	$U(\{\mathit{conflicthistory}, \mathit{vdem}, \mathit{wdi}, \mathit{ged}_{\text{sb}}, \mathit{all}\})$

Table A.1: Distributions the hyperparameters are sampled from for the random search. U denotes the uniform distribution, U_d the discrete uniform distribution and Ber the Bernoulli distribution.

The number of hidden layers h is set to a maximum of 6 to mitigate overfitting and the batch size b is kept relatively small due to limited data availability. Adam is used as the optimization algorithm (Kingma and Ba 2017) with a minimum learning rate of 0.001. In our case, smaller learning rates tend to produce worse results. The feature subset f_{set} is randomly drawn from five possible sets, see Section 2, where ged_{sb} denotes fatality data only and *all* includes all features. We consider lead time $s = 8$ in the tuning process.

Appendix B Figures



(a) 2018

(b) 2019

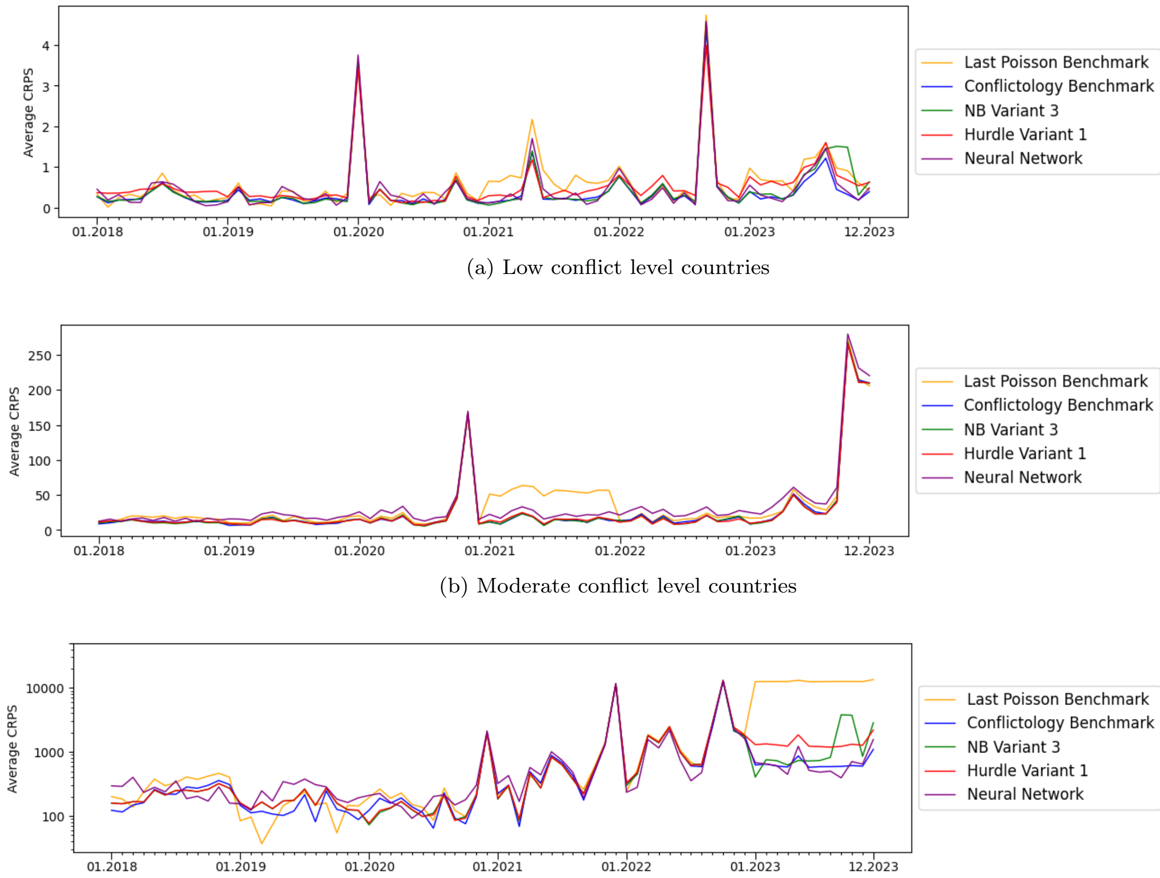
(c) 2020

(d) 2021

(e) 2022

(f) 2023

Figure B.1: Optimal models per country by average CRPS per year within Task 2 (2018–2023). *No preferred model* indicates countries with zero predicted and observed fatalities, since these coincide for all models.



(a) Low conflict level countries

(b) Moderate conflict level countries

(c) High conflict level countries

Figure B.2: Average CRPS per month within Task 2.