Bayesian Density Forecasts for VIEWS*

Patrick T. Brandt
School of Economic, Political, and Policy Sciences
University of Texas, Dallas
pbrandt@utdallas.edu

June 25, 2024

Abstract

Dynamic analysis in VIEWS are important forecasts because the data are inherently serially correlated (in space, time and both). Here we consider dynamic forecasts as a baseline: above the density baselines proposed as part of VIEWS 2.0, the forecasts proposed here provide Bayesian density forecasts that allow for the evaluation of simple dynamics (autoregression and time trends) and for different distributional assumptions (e.g., Poisson, negative binomial, zero-inflated, Tweedie). The idea here is that the forecasts proposed should be baseline in the sense that no-change or density forecasts account for the basic properties of the data. While modest, the idea is that the baseline bar for a "success" here can and should be higher.

*This paper documents a contribution to the VIEWS Prediction Challenge 2023/2024. 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, H. et al. (2024) and https://viewsforecasting.org/research/prediction-challenge-2023. Thanks for help with code and data go to Dhanish Parimalakumar. Thanks also for discussions and feedback from Vito D'Orazio. All errors are solely the responsibility of the author. This material is based upon work supported by the National Science Foundation under Grant Nos. OAC-CSSI-1931541 and 2311142. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Replication code and data are available from the author and via https://github.com/PTB-OEDA/VIEWS2-DensityForecasts.

1 Introduction

This outlines the presentation and specifications of the forecast models contributed to Hegre, H. et al. (2024). The goal here is to take a fully Bayesian approach that admits that a the conflict forecaster is uncertain about the following choices:

1. Data in the sample / training: this in the main is addressed by the design of the forecasting competition itself.

2. Choice of the parametric density (e.g., Poisson, compound Poisson / Tweedie, negative binomial). A review of the papers from the previous event shows that this is a question to be considered and especially in light of the forecast baseline being a simple Poisson distribution (to be estimated based on the training sample chosen). For an overview of the compound Poisson or Tweedie exponential dispersion models used below, see Jorgensen (1997)

3. Specification of the dynamics: autoregression, local trends, global trends, and relative weighting of each. In the earlier analyses of the `pgm` and `cm` data various approaches to *nearness* in space-time were considered for how serially correlated the events predicted may be. The idea here is to benchmark and systematize how one at least thinks about this in the time dimension for hierarchical data (Pedersen et al., 2019).

4. Evaluation of the forecast densities: focus on distributions, not on single quantiles nor on summary statistics (means, medians). This requires thinking about the comparing the relative costs via Murphy diagrams, CRPS, DRPS, Taylor plots, and forecast variances, etc.

Being "Bayesian" in this context is that models and probabilistic statements are open for interpretation and the weighting of evidence about what one more or less sees as true (for approaches to this see Gill (2021) and McElreath (2018)). It is not about (overly) rigorous beliefs about one class or set of prior beliefs about the data, the forecast models, or the parmeters alone (in these models). Based on in- and out-of-sample comparisons an assessment will and ought to be made that looks at the relative properties and beliefs about the performance of models and their forecast densities.

2 Models / methods considered

The goal here is to use a Bayesian method to select the likelihood-forecast density class (Poisson (P), negative binomial (NB), or Tweedie (TW)). Then these densities are applied in different GAM / GLMM modeling contexts to address the dynamics. The GAM and GLMM models (detailed below) are 1) local (country-specific time splines), and 2) GLMM (time and country specific effects with autoregression).[1]

For each forecast period of VIEWS forecasts are generated with the following options:

$$
\begin{aligned}
\text{forecasts set} &= \#densities \times \#methods \times \#horizons(months) \\
&= \begin{pmatrix} \text{Poisson (P)} \\ \text{Negative binomial (NB)} \\ \text{Tweedie (TW)} \end{pmatrix} \times \begin{pmatrix} \text{GAM local} \\ \text{GLMM} \end{pmatrix} \times \begin{pmatrix} 2018:1-2018:12 \\ 2019:1-2019:12 \\ 2020:1-2020:12 \\ 2021:1-2021:12 \\ 2022:1-2022:12 \\ 2023:1-2023:12 \\ 2024:7-2025:6 \end{pmatrix}
\end{aligned}
$$

There are 2 additional forecasts each year to account for the gap between the training data samples defined in the project and the forecast reporting periods.

[1] The previous version of this analysis also include tensor GAM specifications (October 2023). These are not presented here since they do not fit as well as the models reported herein.

2.1 GAM models

The GAMs allow for general interactions of the possible (local) time trends and their interactions and commonalities with the country identifiers.[2] In addition to the time and unit random effects, autoregression terms are include. The GAM here has the form:

$$y_{it} \sim Count(\mu_{it}) \quad \text{where} \quad log(\mu_{it}) = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 f_1(t_i) + \beta_{3i} f_2(v_i, t_i). \tag{1}$$

A key Bayesian choice of the selection of the parametric distribution for the observed counts y_{it} (`ged_sb` in the analysis).

The choices for the $Count(\cdot)$ distribution are

1. Poisson

2. Negative binomial (with estimation of the variance scale parameter θ).

3. Tweedie (with estimation of the scale power parameter p).

The function $f_1(t_i)$ are the time-splines for the country i. The $f_2(t_i, v_i)$ are a tensor-grid basis for country and time effects. In the GAM estimates are produced with $\beta_3 = 0$, so only lag dynamics and temporal splines are included. Allowing $\beta_3 \neq 0$ is the earlier "tensor" specification (not reported here).

2.2 GLMM models

The GLMM specifications follow the standard form such that for the counts of `sb_ged` $= y_{it}$ for country i in month t:

$$g\left(\mathrm{E}[y|u]\right) = X\beta + Zu \tag{2}$$

[2] Note, that computationally this is feasible for the `cm` datasets. But it is not likely to be possible for the `pgm` data. These are done using the canonical R package `mgcv`

where $g()$ is the canonical link function, X are the linear predictor (here just a common intercept) and Z are the group-fixed effects.[3] The Z include autoregressive dynamics and monthly trends for each country.

3 Forecast description and formats

Sample scale target: `cm`

Initial training sample: Through October 2017.

Rolling / update decision: Models are updated for each forecast horizon. For example, the 2020 forecasts include all prior data (origin to 2019(10)).

Model and File naming convention: In the folder submission structure for VIEWS, the naming convention is

`/count_model/cm/test_window=YYYY/count_model_YYYY.parquet`

Here the file and path notation is

count is `poisson, neg_binom, tweedie`,

model is `gam, glmm`,

YYYY is `2018, 2019, 2020, 2021, 2022, 2023, 2024`.

As requested there are separate `*.yml` files in each `/count_model/` folder for the forecast model descriptions.

4 Preliminary findings and conclusions

In-sample information criteria (AIC) in Table 1 support the choices of models with explicit dispersions terms (so for GAM and GLMM, the negative binomial and Tweedie models).

[3] Implementation details are at `https://github.com/glmmTMB/glmmTMB`.

	GAM			GLMM		
Training End	Poisson	NegBin	Tweedie	Poisson	NegBin	Tweedie
2017	Inf	95701	95534	94278	91710	91674
2018	Inf	99607	99412	98196	95459	95413
2019	1593986	103579	103386	102125	99245	99191
2020	1969312	107636	107439	106112		103035
2021	Inf	111908	111903	110296		107127
2022	Inf	116543	116893	114655	111465	111362
2023	Inf	121051	121575	118856	115571	115447

Table 1: AIC values for different training sets and models

Once out-of-sample forecast performance is assessed via CRPS metrics, results become more mixed (as expected), but generally support the main point that *simple benchmarking can go beyond just draws from a Poisson: in fact this is too low of a bar and too many models can easily beat it.* Table 2 shows the CRPS metrics for the models described above and comparisons to the various VIEWS benchmarks in the final 3 columns. GLMM models with densities that allow for over-dipersion do best for forecasts from 2018–2021. They are nearly identical to the various VIEWS baseline models in 2022. The 2023 data diverge widely and the forecast models' performances are generally quite poor. This is due to the large changes in the mean, variance, and upper quantiles of the data since 2021 — as a consquence of the Tigray, Ukraine-Russia, and Israel conflicts.[4]

Figure 1 shows the CRPS by month for each model. Note that they are all quite similar from 2018:1–2021:6, growing much more volatile after 2020:6.

Figure 2 has two rows. The first row are the monthly mean counts for the indicated months and years. The second row gives the ratio of the predicted standard deviation for each of the forecast models in Table 2 to the observed data standard deviations by month. Such a ratio of the model to the data standard deviations as a forecast metric is suggested by Dietze (2017) to check for temporal forecast breakdowns. Ratios near one indicate that the forecasts are near the data and deviations above (below) indicate forecast variance greater (less) than the data variation. The ratios indicate good overall forecast performance for the

[4]Additional tables and figures in the Appendix below show this.

Forecast Year	GAM			GLMM				Constituent	Exactly
	Poisson	Neg.Binom	Tweedie	Poisson	Neg Binom	Tweedie	Bootstrap	Poisson	Zero
2018	38832015.90	7546550.45	21.55	**16.31**	*16.69*	17.06	23.58	20.17	24.13
2019	54104307.13	7360561.70	11.62	12.30	**8.61**	*8.76*	22.46	9.48	23.02
2020	33228633.54	6736348.29	25.29	22.25	**21.51**	*21.97*	31.42	23.70	32.04
2021	38394555.18	4800238.95	*82.51*	84.10	**82.45**	83.06	86.63	85.61	87.34
2022	41901335.72	15706903.49	3459661.85	124.76	127.04	127.66	**120.25**	131.02	*120.97*
2023	59410151.01	14612157.73	6054067.71	409.89	536.02	591.87	**52.72**	678.96	*53.54*

Table 2: CRPS by forecast year based on all prior periods as training data. Smallest CRPS values for each year noted in **bold**, second smallest in *italics*.

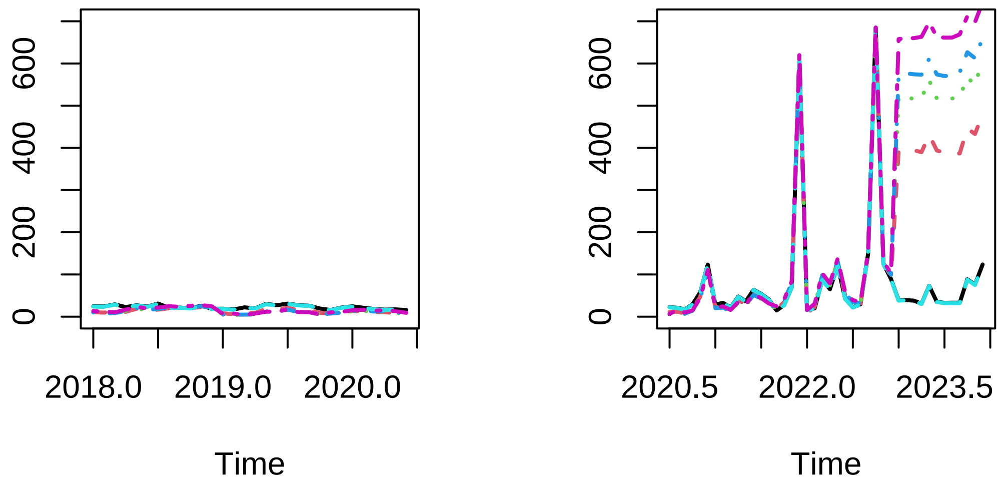

Figure 1: CRPS by model, by month. Each model is a different color / line type as given by columns in Table 2.

baseline models (since they reproduced the data standard deviations on a static basis for the rolling training data). Performance for the GLMM models forecast standard deviations vary from 2018–2022 with some models — GLMM Poisson [black], GLMM negative binomial [red], and GLMM Tweedie [green] below and above one. All of forecast models proposed here (including non-baseline / benchmarks) overstate the forecast variance relative to the data in 2023 (bottom right panel).

Finally, Figure 3 implements the plot of Taylor (2001). This figure places the normalized standard deviation of the data on the x-axis and compares it to the normalized standard deviation of the forecasts on the y-axis. This can then be conditioned over the posterior forecast sample attributes (models, time periods, etc.). Based on the law of cosines these quantities can be used to relate the correlation and the standard deviation of the data and

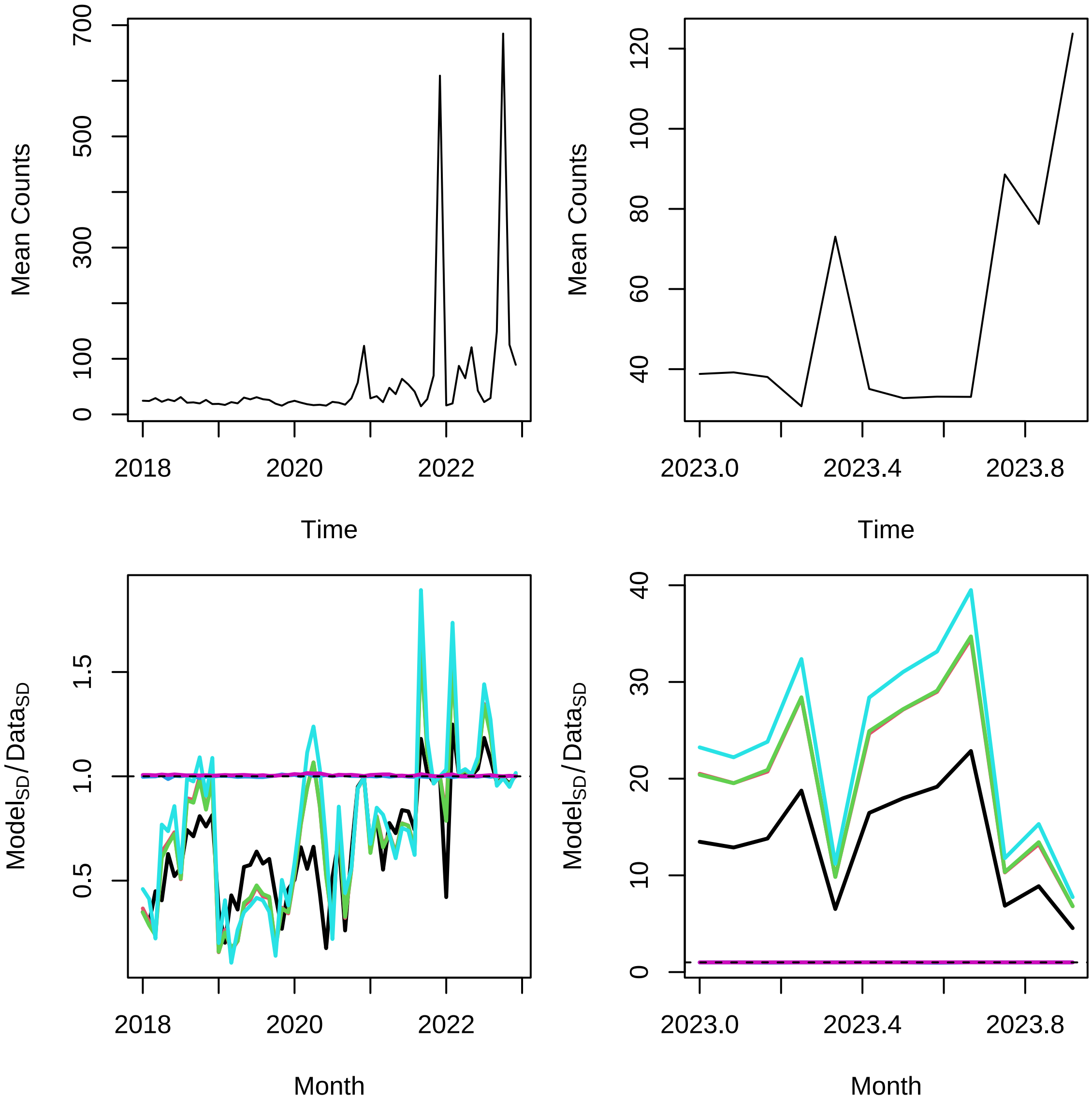

Figure 2: Mean event counts (row 1) and Forecast model standard deviations to data standard deviations ratios (row 2), by month. Each model is a different color / line type as given in Table 2.

the forecasts.[5] "Best forecasts" would be highly correlated and share the same normalized value with the data at $(1,0)$ in the plot. Conditional (model) forecasts below this point and toward the origin, have too little variance relative to the data. For 2018 and 2019 the GLMM models are well correlated with the observed events and have variance ratios near the observed data (consistent with the earlier ratio plots in Figure 3). But this performance degrades in 2021 and 2022. A (normalized) Taylor plot for 2023 is not included since the extreme values (noted in the Appendix) make the other comparisons impossible to see after normalization.

[5] For two variables X and Y, the centered root mean squared error (CRMSE) is

$$CRMSE(X,Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(x_i-\mu_x)(y_i-\mu_y)]^2} \quad (3)$$

$$CRMSE(X,Y)^2 = \sigma_X^2+\sigma_Y^2-2\sigma_X\sigma_Y R_{XY} \quad (4)$$

where R_{XY} is the correlation of X and Y. Then using the law of cosines ($c^2 = a^2+b^2-2ab\cdot cos(\theta)$), $\theta = arccos(R_{XY})$. For details on implementation, see Anžel et al. (2023); Carslaw and Ropkins (2012).

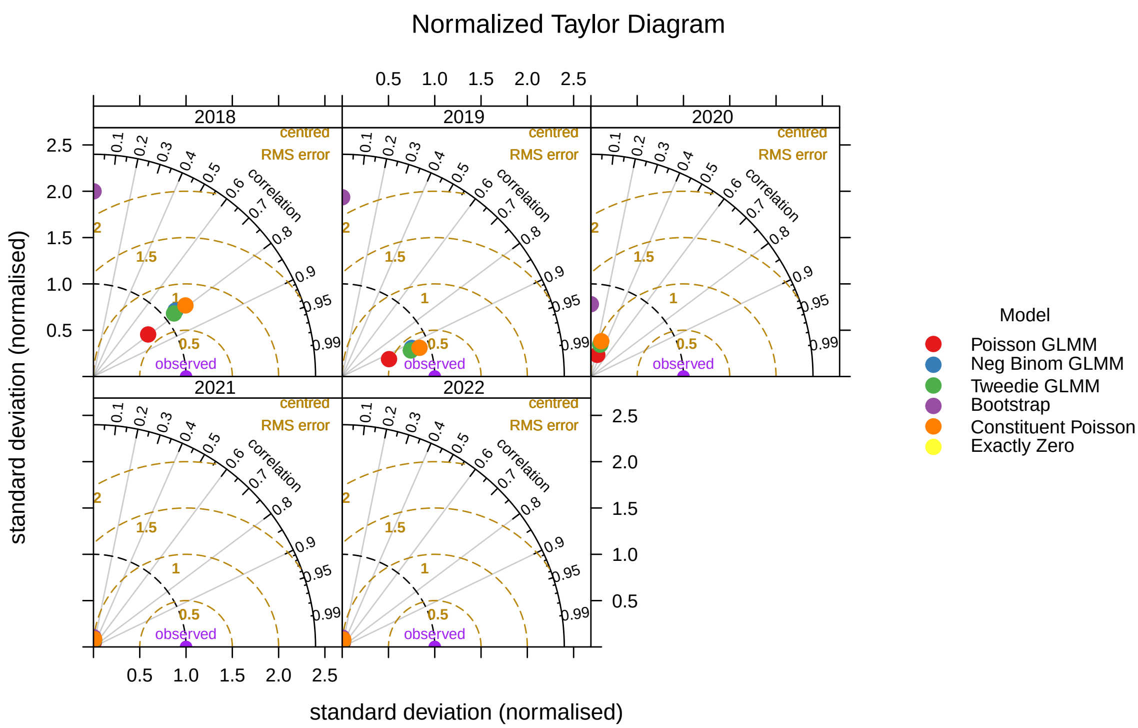

Figure 3: Taylor plot

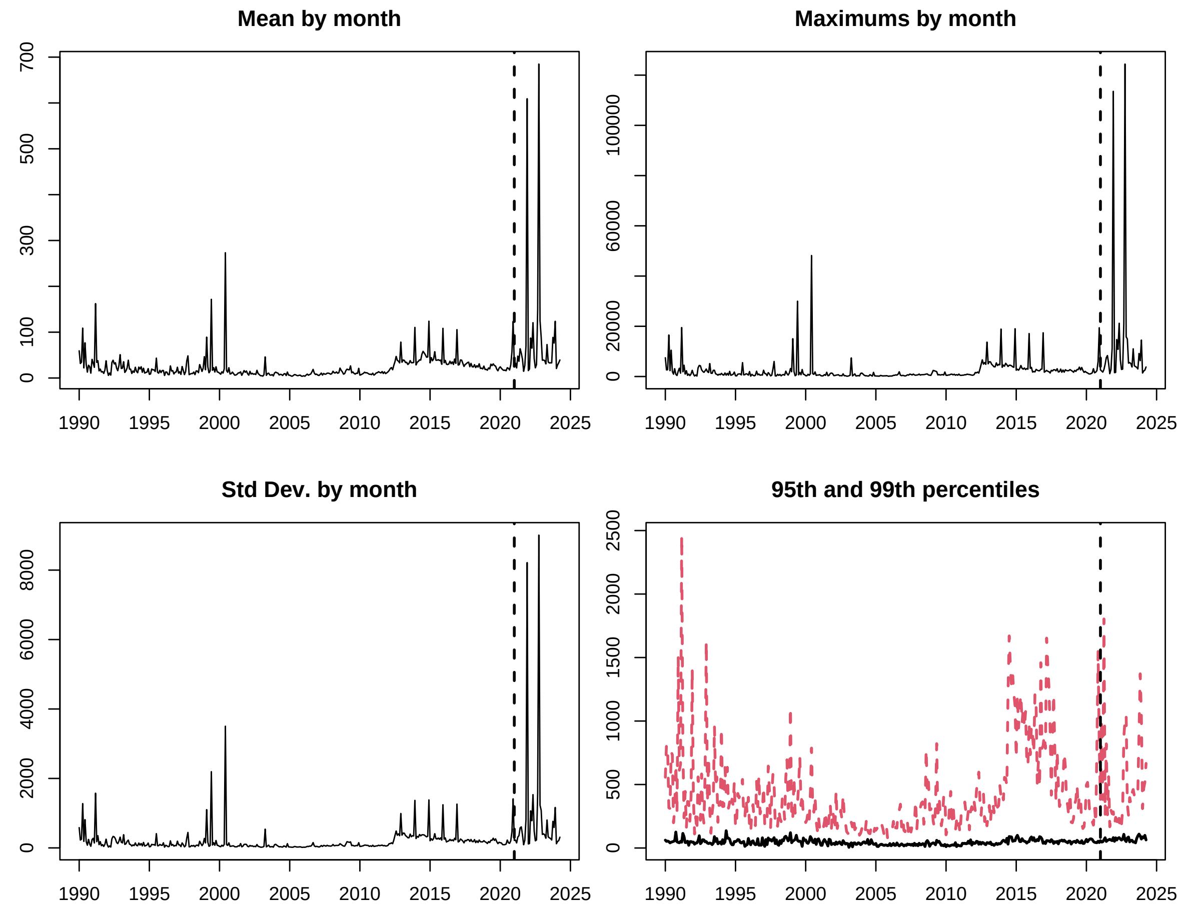

Figure 4: Various monthly summary statistics for the VIEWS `cm` data, 1990-2024. Vertical line is December 2021. Note the y-axes differences.

Appendix: supplemental summary figures and statistics

Table 3: Worst top (99.5%) of the data since 2018, sorted by date

	GED_SB	Country	Date
1	3115	Afghanistan	July 2019
2	2335	Afghanistan	October 2019
3	5991	Azerbaijan	November 2019
4	2247	Afghanistan	November 2019
5	19424	Ethiopia	December 2019
6	2051	Afghanistan	December 2019
7	2657	Afghanistan	January 2020
8	2215	Ethiopia	February 2020
9	2048	Afghanistan	February 2020
10	2793	Yemen	April 2020
11	2768	Afghanistan	April 2020
12	4861	Afghanistan	May 2020
13	2668	Ethiopia	June 2020
14	7512	Afghanistan	June 2020
15	8356	Afghanistan	July 2020
16	5163	Afghanistan	August 2020
17	2782	Yemen	October 2020
18	11246	Yemen	November 2020
19	113554	Ethiopia	December 2020
20	14748	Ukraine	March 2021
21	10810	Ukraine	April 2021
22	21216	Ukraine	May 2021
23	6672	Ukraine	June 2021
24	2804	Ukraine	July 2021
25	2937	Ukraine	August 2021
26	21171	Ethiopia	September 2021
27	4828	Ukraine	September 2021
28	124427	Ethiopia	October 2021
29	4594	Ukraine	October 2021
30	15771	Ethiopia	November 2021
31	6475	Ukraine	November 2021
32	14960	Ukraine	December 2021
33	5310	Ukraine	January 2022
34	5554	Ukraine	February 2022
35	5187	Ukraine	March 2022
36	3782	Ukraine	April 2022
37	11031	Ukraine	May 2022
38	4298	Ukraine	June 2022
39	3939	Ukraine	July 2022
40	3663	Ukraine	August 2022
41	2997	Ukraine	September 2022
42	5296	Ukraine	October 2022
43	9175	Israel	October 2022
44	5034	Ukraine	November 2022
45	6303	Israel	November 2022
46	14514	Ukraine	December 2022
47	7068	Israel	December 2022
48	2197	Ukraine	February 2023
49	2551	Israel	March 2023
50	3804	Ukraine	April 2023

References

Anžel, A., D. Heider, and G. Hattab (2023). Interactive polar diagrams for model comparison. *Computer Methods and Programs in Biomedicine 242*, 107843.

Carslaw, D. C. and K. Ropkins (2012). openair — An R package for air quality data analysis. *Environmental Modelling & Software 27–28*(0), 52–61.

Dietze, M. (2017). *Ecological forecasting*. Princeton University Press.

Gill, J. (2021). Political science is a data science. *The Journal of Politics 83*(1), 1–7.

Hegre, H. et al. (2024). The 2023/24 VIEWS prediction competition. *Journal of Peace Research*.

Jorgensen, B. (1997). *The theory of dispersion models*. Chapman & Hall / CRC.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ 7*, e6876.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres 106*(D7), 7183–7192.