Retrospective evaluation; forecasts July 2018 – December 2019.

Online appendix E to ViEWS$_{2020}$: Revising and evaluating the ViEWS political Violence Early-Warning System.

Håvard Hegre[1,2], Curtis Bell[1,3], Michael Colaresi[1,4], Mihai Croicu[1], Frederick Hoyles[1], Remco Jansen[1], Maxine Ria Leis[1], Angelica Lindqvist-McGowan[1], David Randahl[1], Espen Geelmuyden Rød[1], and Paola Vesco[1]

[1]Department of Peace and Conflict Research, Uppsala University
[2]Peace Research Institute Oslo
[3]U.S. Naval War College
[4]University of Pittsburgh

January 22, 2021

CONTENTS

E-1   COMPARING PREVIOUS PUBLISHED FORECASTS WITH ACTUAL EVENTS

In this appendix, we assess how well the ViEWS system in place up to February 2020 was performing against what actually happened. For details on the setup of ViEWS, definitions of dependent variables and units of analysis as well as an explanation of the notation we use, see Appendix A – ViEWS: Data setup and notation. For evaluation of the current ViEWS system and a comparison with the former system, see Appendix D – Evaluation of models and ensembles.

ViEWS has produced updated forecasts every month since July 2018 based on the setup documented in Hegre et al. (2019). We will refer to the individual forecast runs as r_2018_07, r_2018_08, et cetera below. The model setup changed slightly from r_2018_09 to r_2018_10, but has been kept similar since then. We use actual conflict data from two sources to evaluate the forecasting results: i) UCDP-GED (Pettersson, Högbladh, and Öberg, 2019) up to December 2018, and ii) UCDP-Candidate (Hegre et al., 2018) thereafter. Since the UCDP-GED data have been vetted more carefully than UCDP-Candidate, some systematic differences between the two periods may be reflected in the evaluation. The UCDP-Candidate dataset for 2019 also include a small number of coding errors that will be corrected before entering the final dataset.

Figure E-1. Areas under the Precision-Recall curve (AUPR, left), Area under the Receiver Operator Curve (AUROC, right), averaged across all runs with predictions, by $step$ (number of months ahead)[a]
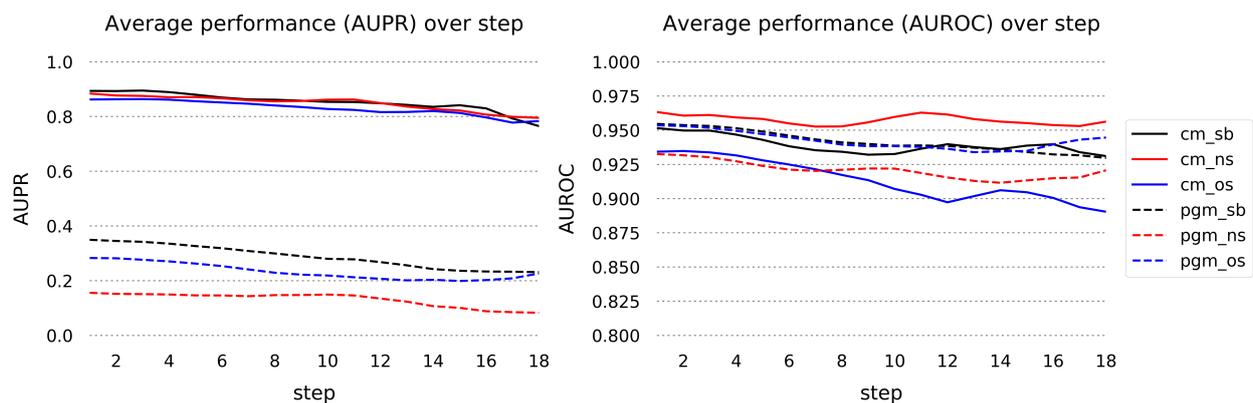
[a]All series are 3-month moving averages. For $s = 1$, the plotted point is the average for all runs from July 2018 through December 2019. The number of runs to average over decreases gradually. For $s = 18$, we only have predictions from the July 2018 run to evaluate.

Figure E-1 shows the areas under the precision-recall curves (left) and the area under the receiver-operator curve (right) by s, i.e. the number of steps/months into the future for which we have produced a forecast, for each outcome **sb**, **os**, **ns**, and each level of analysis.[1] The results shown in the figure largely corresponds with what we found out-of-sample from our test window (2015–2017) in Hegre et al. (2019, Figures 4, S-10, S-11).[2] The predictive performance of the ensembles is excellent, and not far from what we expected. For the *cm* outcomes, AUPR is about 0.9 for all conflict types when forecasting a few months into the future, and decreases to about 0.8 when $s \geq 16$. The Area under the Receiver Operator Curve (AUROC) decreases from about 0.95 when s is small to 0.925 when $s \geq 12$. This performance is somewhat poorer than what we indicated in the out-of-sample evaluation we published in Hegre et al. (2019).

The corresponding metrics for the *pgm* level are closer to the performance in the out-of-sample evaluation of last year. The AUPR for state-based conflict (**sb**) is 0.3 for low s and decreases to 0.2 when forecasting 12–13 months into the future. The true performance of ViEWS for non-state conflict is additionally slightly better than expected, with a relatively persistent score at around 0.15 instead of 0.10 as we found out-of-

---

[1]We can evaluate forecasts one month into the future using predictions from all the 18 published forecasts. For $s = 18$, we only have forecasts from the July 2018 run.

[2]Figures S-10 and S-11 are found in the online appendix, available at `https://pcr.uu.se/research/views/downloads/jpr-2019-2-material/`.

sample over 2015-2017. The AUROC starts at about 0.925 and decreases to about 0.915 at $s = 12$ as in the original out-of-sample evaluation Hegre et al. (2019).

E-2   FOCUSING ON DECEMBER 2019

We will look at the forecasts for December 2019 in some more detail. Figure E-2, E-3, and E-4 show where UCDP-Candidate recorded **sb**, **ns**, and **os** conflict events in that month, along with the corresponding ensemble forecasts we published in July 2019 for $s = 6$ months into the future at the *pgm* (PRIO-GRID month) level. The maps indicate that the ensemble models did quite well at forecasting the locations of actual events. For **sb**, only a cluster of events in central Mozambique was unanticipated, aside from other incidental event locations.

At the *cm* level, the model criticism plots in Figure E-5 are informative. Ordering the individual country-month forecasts by the value of predicted probability, these plots show the ten units that are the furthest away from 0 or 1, categorized by actual negatives (blue) and actual positives (red) respectively. They also visualize how well the models separate in terms of the probability of the outcome. The models separate moderately well. The probability distributions for the actual negative cases are centered quite densely between 0 and 0.25, with a lengthy right tail. The probability distributions for the actual positives are also quite dense, but centered around probabilities of 0.5.

Given that actual positives are sparse on a monthly scale, the figures are most informative with regard to identifying potential false positives.[3] For December 2019 on the outcome of **sb**, Sudan and Ethiopia were the worst misses, with a forecast value between 0.4 and 0.5 at $s = 6$, and with Sudan even on 0.6 at $s = 12$. These are historically high-risk countries, but have been becoming more stable over recent months. Among the worst potential false negatives on the outcome of **sb**, Mozambique and Burkina Faso (BFA) are of note. Historically not as violent, the recent escalation of Jihadist violence in both these countries did not translate into a commensurate rise in the predicted probability of **sb** violence in our model ensemble, although it did increase between the January 2019 and July 2019 runs. Regarding the outcome of **ns**, Libya, Kenya, and South Sudan were off the most, both at $s = 6$ and $s = 12$. The two strongest potential false negatives here for December 2019 are South Africa and Cameroon. In South Africa, an incident of gang violence was recorded in the UCDP candidate data, while for Cameroon, non-state violence had not been very frequent prior to the emergence of the Anglophone separatist movement in September 2017. For **os**, Ethiopia and Kenya were the furthest away from 0, with a forecast value at around 0.5 for both $s = 6$ and $s = 12$. The figures here also show two strong false negatives: Malawi and Namibia. The case of Malawi is a coding error in UCDP-candidate that will be corrected when UCDP-GED for 2019 is released. The case of Namibia is one of police violence initially included in the candidate data, but later set to 'unclear' (and thus excluded from the data informing the ViEWS forecasts) due to conflicting information about the incident.

Figure E-6 shows the model criticism plot for each run between July 2018 and December 2019, per $s = 1$ and $s = 6$. Note that $s = 6$ has fewer observations, as the last predictions with actuals to compare to in that case were produced five months prior, in 2019-07.

Figure E-8 shows how the AUPR of our forecasts for December 2019 changed from our first published models 18 months earlier, to the one we produced in late November 2019. The model ensemble steadily produced better **sb** and **os** predictions at both the *cm* and *pgm* level in the run up to December 2019 as we accumulated new conflict data. The performance for the **ns** predictions has remained relatively stable, reflecting that these events are more rare and display less systematic patterns than **sb** and **os** events.

Figure E-7 shows how our forecasts react to events. Since events data are produced with a one-month lag, new events can only affect the forecasts for two months later. In March 2019, for instance, UCDP recorded

---

[3]Note 'potential': whether or not individual forecasts are false positives depends on the classification threshold.

Figure E-2. Map of *pgm*-**sb** forecasts for December 2019 from r_2019_07, ensemble model, with actuals
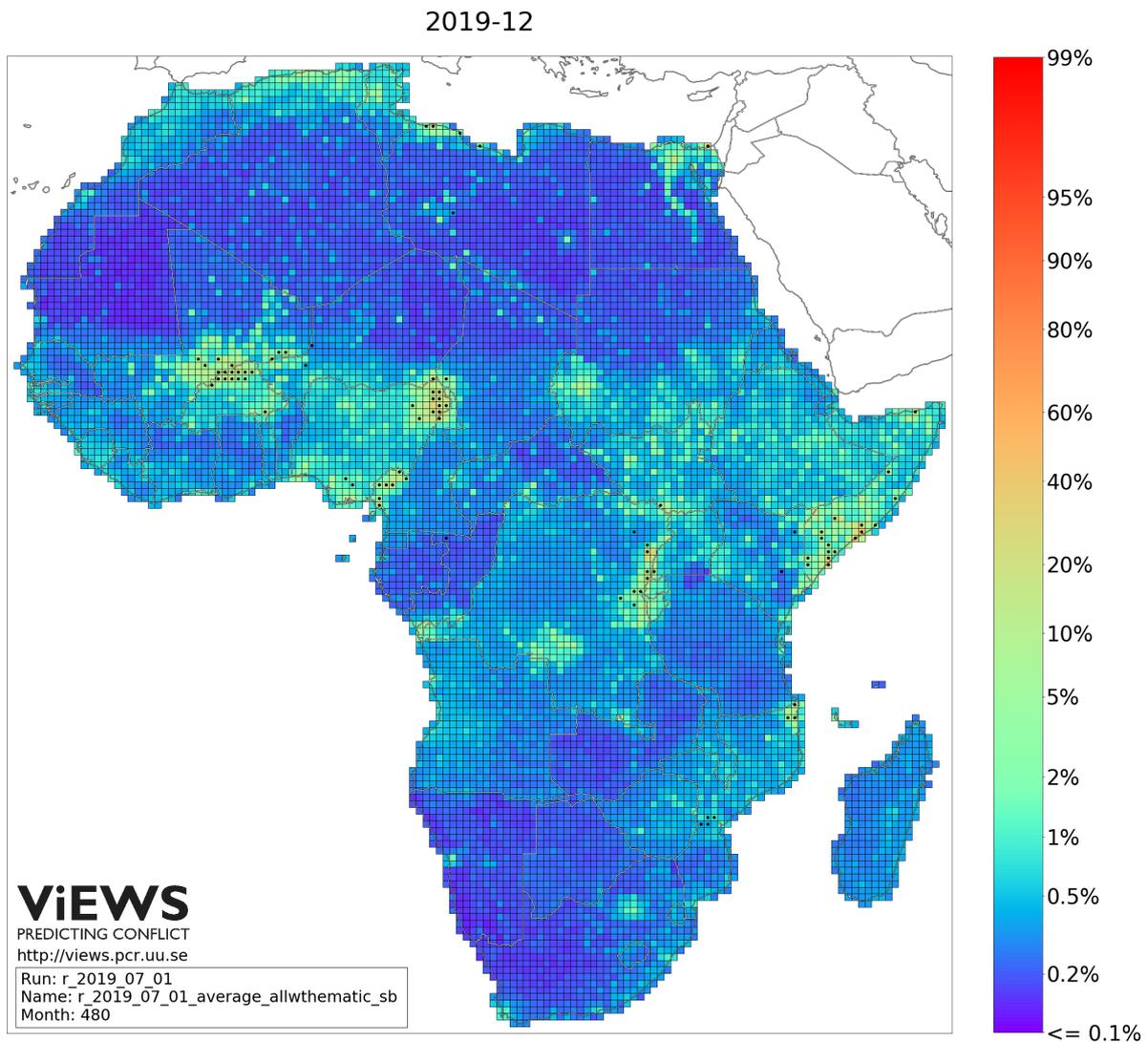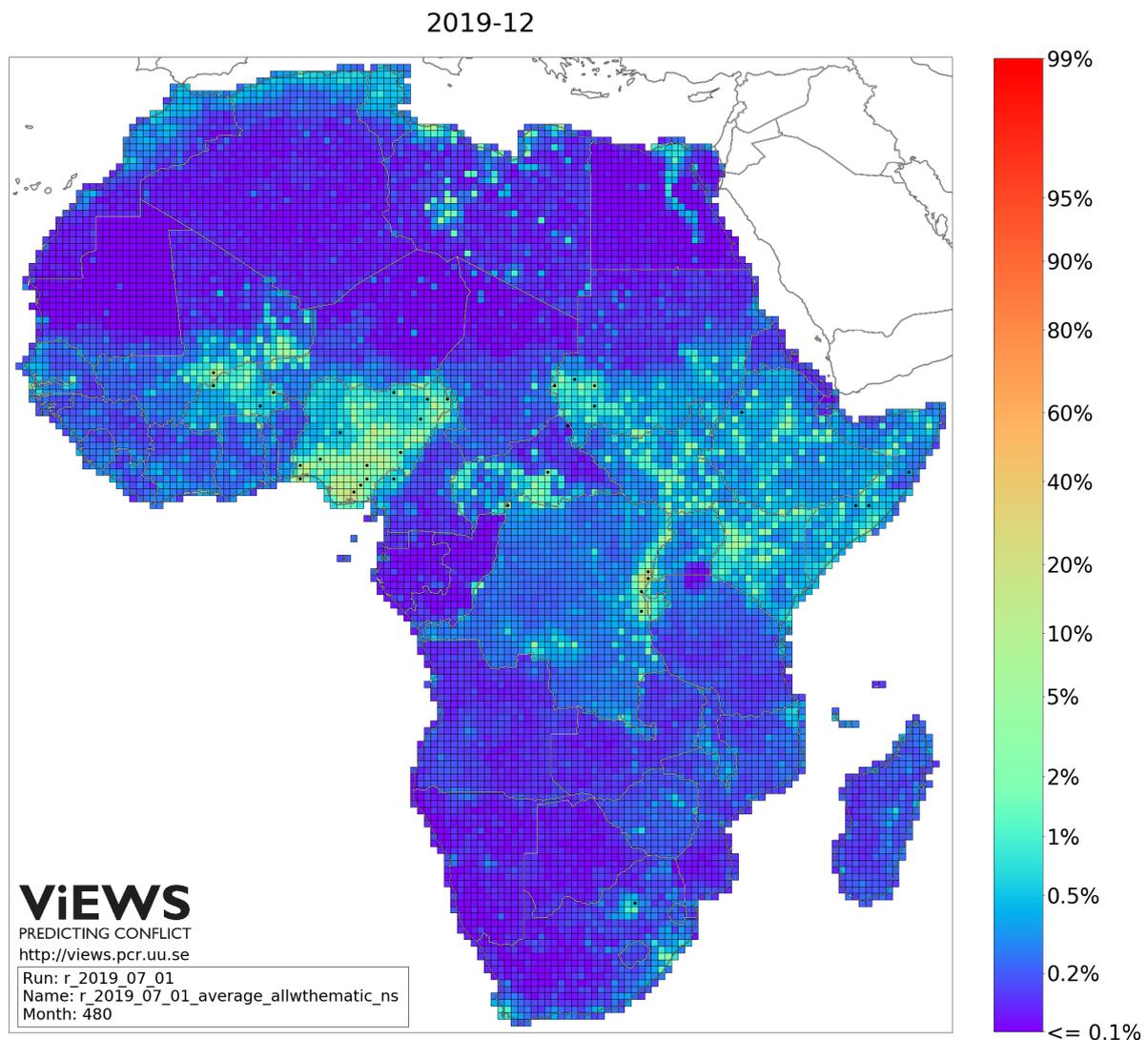

2019-12

Note: The colors in the map show the predicted probabilities of an event with at least one battle-related death classified as state-based conflict within each PRIO-GRID cell for December 2019. The predictions were generated in early July 2019 (or r_2019_07 run). The black dots show where the UCDP candidate event dataset recorded such conflict in their January 2020 release. The long-standing conflicts in Mali, North Nigeria, the Kivu provinces, Somalia, and Libya account for the majority of the cases. The somewhat more recent conflicts in Burkina Faso, Cameroon, Ituri province, and North Mozambique account for most of the remainders. Additionally, there has emerged an entirely new cluster of conflict events in central Mozambique, where a splinter group from RENAMO initiated fighting against the government.

a **sb** event in the Comoros. This is reflected in our forecasts for May 2019 that show a clear increase in the predicted proportion of pgm **sb** events after.[4]

---

[4]ViEWS changed the composition of the ensemble from run r_2018_09 and r_2018_10. The visible change in the figures from September to November 2018 is due to this.

Figure E-3. Map of *pgm*-**ns** forecasts for December 2019 from r_2019_07, ensemble model, with actuals

2019-12

Note: The colors in the map show the predicted probabilities of an event with at least one battle-related death classified as non-state conflict within each PRIO-GRID cell for December 2019. The predictions were generated in early July 2019 (or r_2019_07 run). The black dots show where the UCDP candidate event dataset recorded such conflict in their January 2020 release. The long-standing conflicts in Mali, Nigeria, Central African Republic, the Kivu provinces, Somalia, and Darfur account for the majority of the cases.

Figure E-4. Map of *pgm*-**os** forecasts for December 2019 from r_2019_07, ensemble model, with actuals
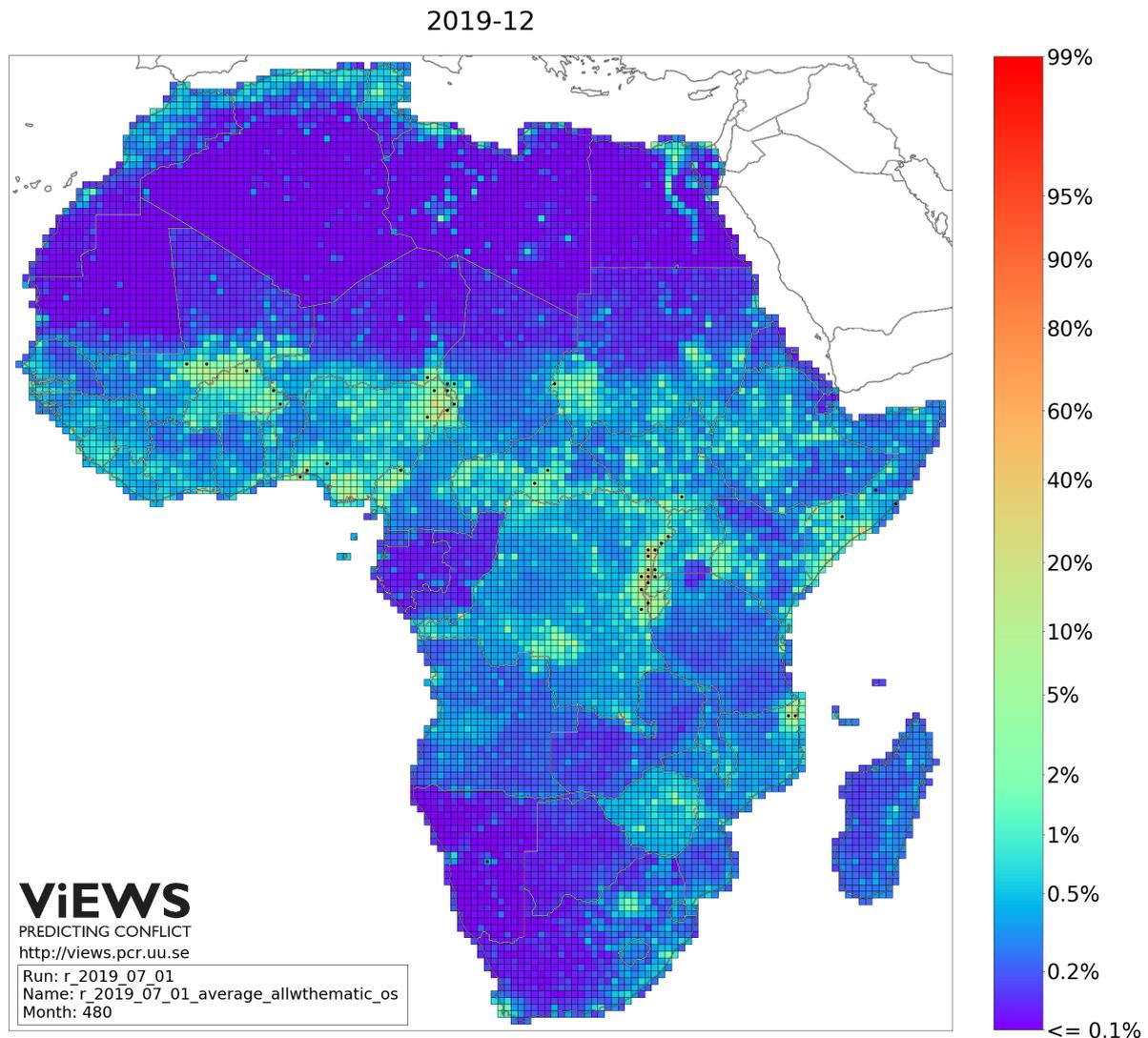
Note: The colors in the map show the predicted probabilities of an event with at least one battle-related death classified as one-sided violence within each PRIO-GRID cell for December 2019. The predictions were generated in early July 2019 (or r_2019_07 run). The black dots show where the UCDP candidate event dataset recorded such conflict in their January 2020 release. Violence mostly on the account of insurgents in Mali, Burkina Faso, northeastern Nigeria, the Kivu provinces, Somalia, northeastern Mozambique account for the majority of the cases.

Figure E-5. Model criticism plots per outcome for December 2019 based on forecasts from
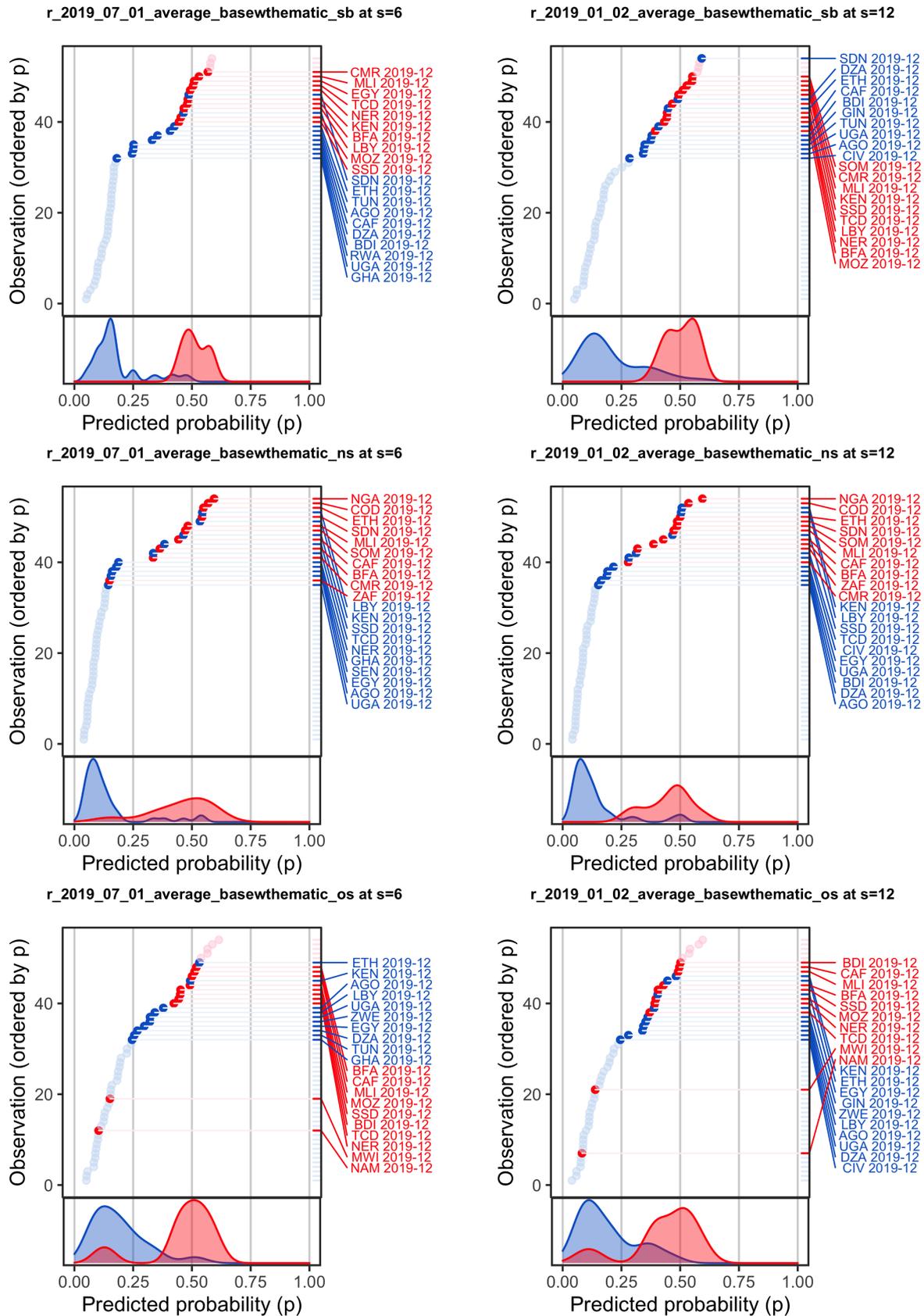r_2019_07 (s=6, left), and r_2019_01 (s=12, right)

Figure E-6. Model criticism plots over all runs, per step $s = 1$ (left), and $s = 6$ (right)
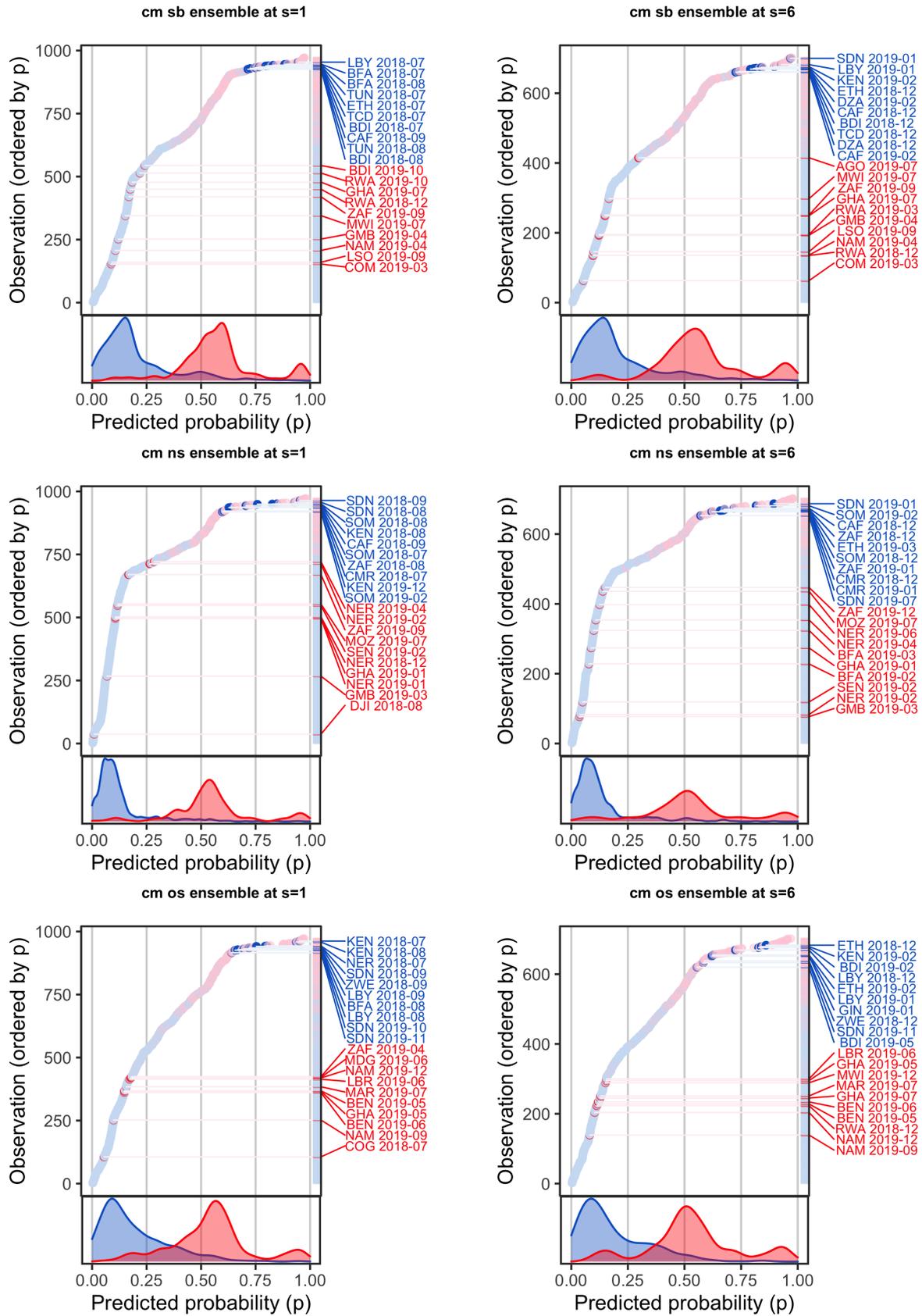
Figure E-7. Predicted probabilities per violence type at country-month (left), and log predicted proportions of PRIO-GRID cells per violence type at country-month (right)
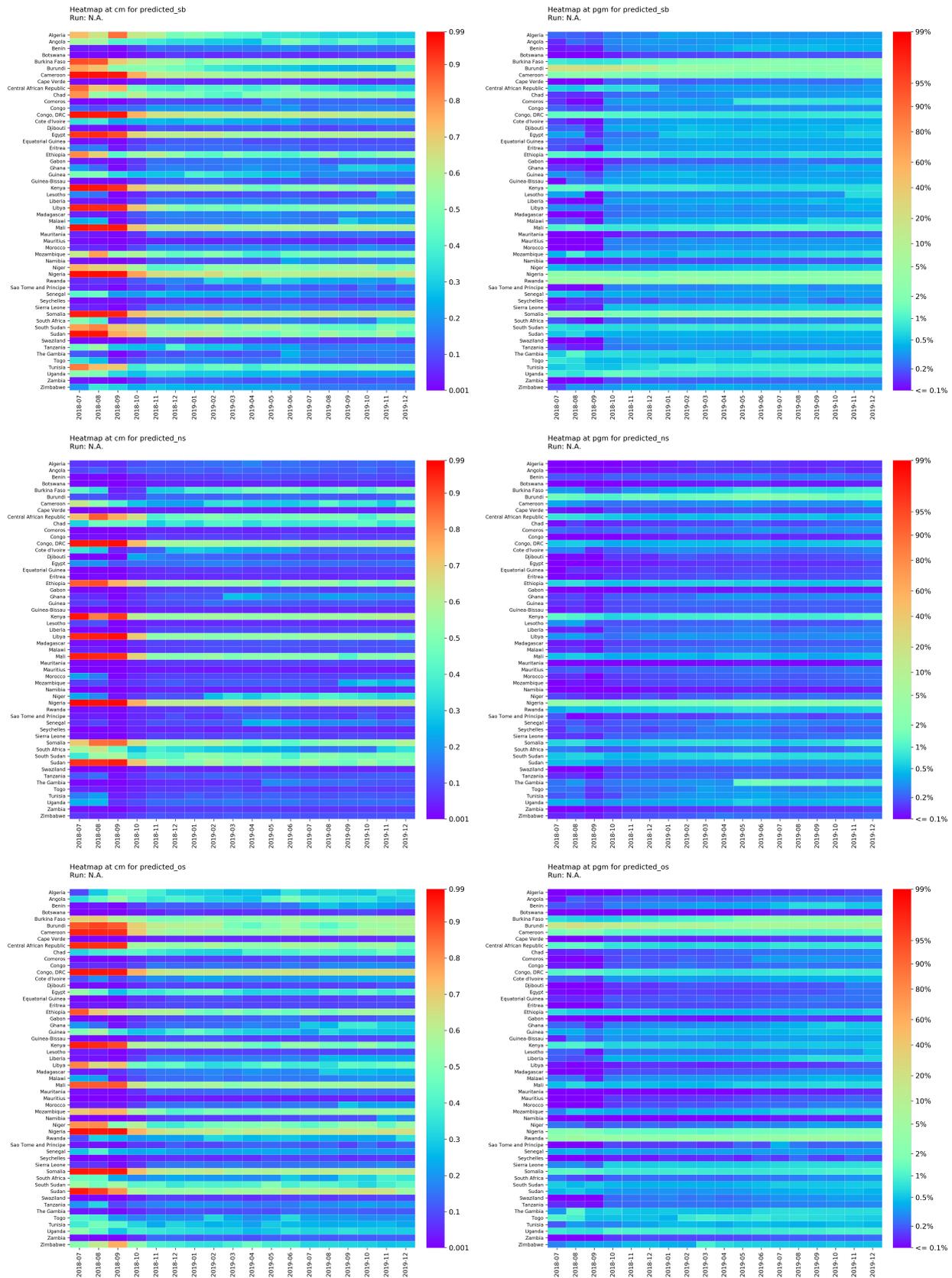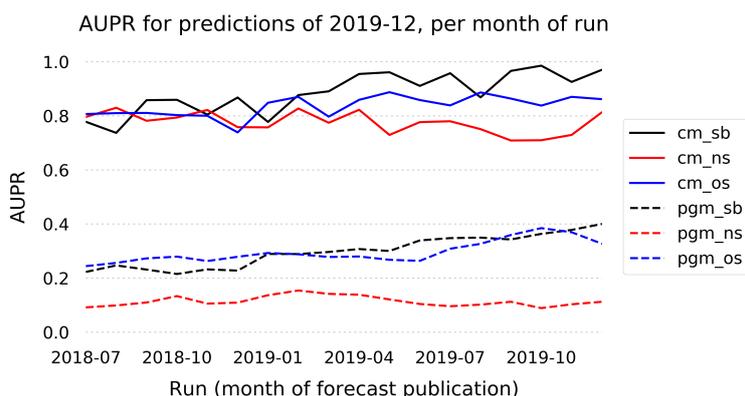
Figure E-8. Area under the precision-recall curve (AUPR) for predictions for July 2019, by month of forecast (run id)

AUPR for predictions of 2019-12, per month of run

E-2.1   A closer look at the Burkina Faso case

Burkina Faso is an interesting case for gaining deeper insight into the forecasting performance and its potential contribution. Up to 2016, the UCDP had not recorded any organized violence events in the country, but at the time of our first published forecasts – the summer of 2018 – and onward, violence escalated considerably in the country. The main actors involved are *JNIM*, a group originally operating in Mali, and the Burkinabe group known as *Ansaroul Islam*. Both groups hold Islamist political agendas and fight the government of Burkina Faso, which according to UCDP (2019) and Human Rights Watch (2019) has been responsible for most of the country's civilian fatalities. To what extent has our model been able to anticipate this escalation?

Figure E-9 shows predictions for $s = 1$ and observed trends for *cm* (left) and *pgm* (right) in Burkina Faso since January 2018.[5] Over the first few months of 2018, a few scattered months with at least one **sb** and **os** conflict event were observed, shown as vertical bars at the *cm* level. The ViEWS *cm* model yielded a high $s = 1$ probability of **sb** and **os** from July 2018 onward. As displayed in the figure, probabilities did not escalate, despite repeated months with conflict events. Given the persistent **sb** and **os** conflicts, the model did however suggest a high probability of **ns** conflict from the summer of 2018 onward. These forecasts turned out to be well-placed, as fatal events between the Islamist groups and the Koglweogo, a Burkinabé self-defense militia, began to occur in the beginning of 2019.

At the *pgm* level, Figure E-9 shows the observed and predicted proportions of PRIO-GRID cells experiencing conflict for each month, and Figure E-10 shows the predicted probabilities as well as the locations of violence per PRIO-GRID cell. From September 2018 onward, violence spread from the northern to the eastern and south-eastern regions. The *pgm* **sb** model responds by a slightly increasing predicted proportion of the country. Figure E-10 shows that the model adapts and does reasonably well at predicting the within-conflict location of the new violence.

Surprisingly, the rise in one-sided violence is mostly on the account of the government of Burkina Faso, with 40 one-sided conflict events resulting in 342 casualties as recorded by the UCDP (Uppsala Conflict Data Program [UCDP], 2019). According to Human Rights Watch, many civilian victims accused of supporting or harbouring members of the armed Islamic groups have subsequently been killed if they did not provide intelligence about the presence of these armed groups to the government's security forces (Human Rights Watch, 2019). Second to the government, most one-sided violence events have been perpetrated by JNIM, particularly in the country's northern regions.

---

[5]For more details on conflicts in Burkina Faso, see `https://ucdp.uu.se/country/439`.

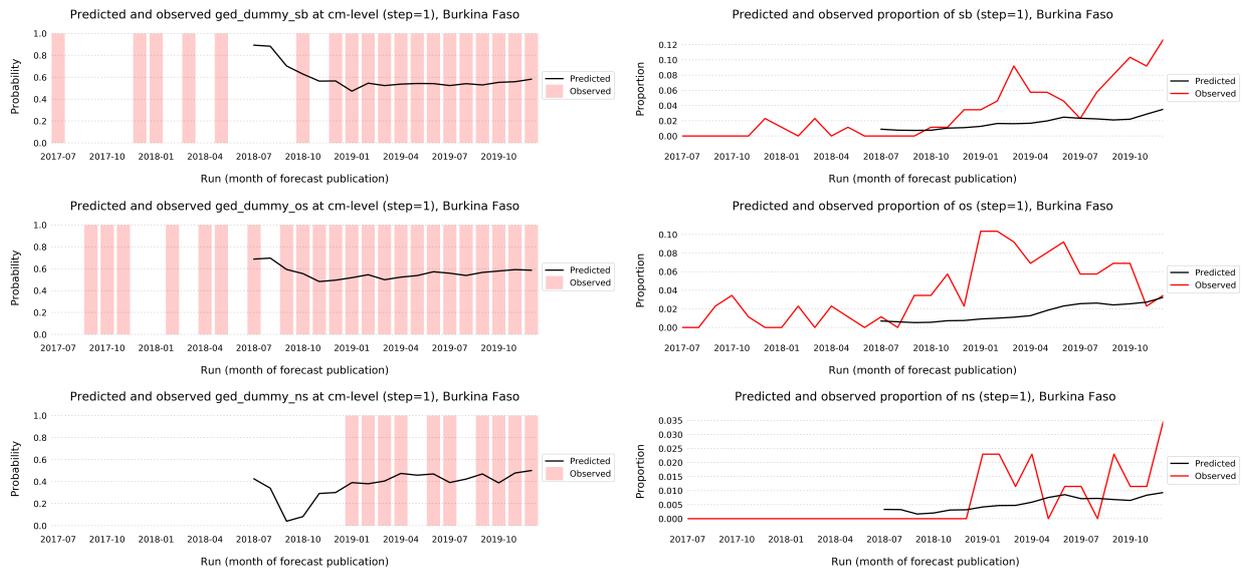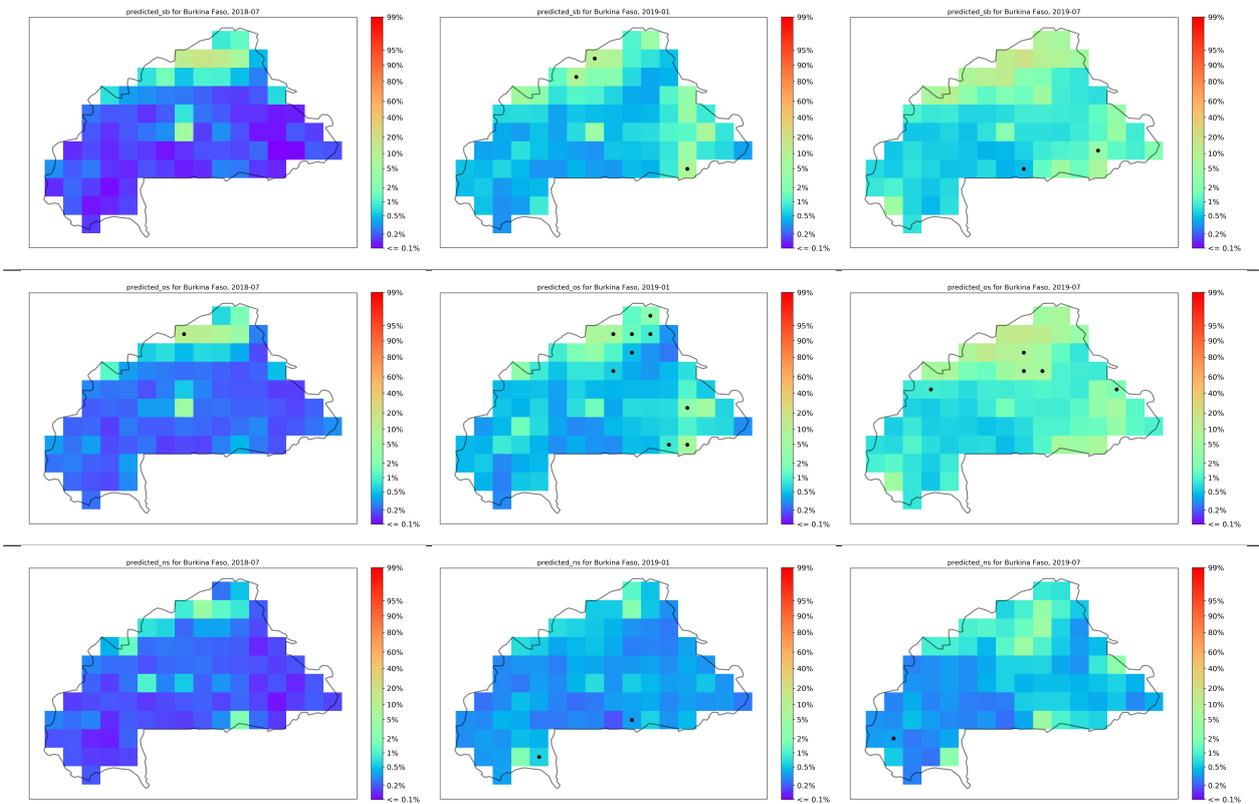*Figure E-9.* Predicted and observed violence at *cm* and *pgm* levels, $s = 1$, Burkina Faso

Figure E-10. Zoom-in of $s = 1$ forecasts for Burkina Faso, July 2018 (left), January 2019 (middle), July 2019 (right), for **sb** (top), **os** (middle), and **ns** (bottom). Markers indicate the actual UCDP Candidate events for the month depicted.

E-2.2   How have our forecasts changed over time?

Figure E-11. Percentage change in forecasted probability of **sb** conflict at *cm* (left) and percentage change in forecasted PRIO-GRID cell proportions (right), ensemble at $s = 1$, August 2018–December 2019
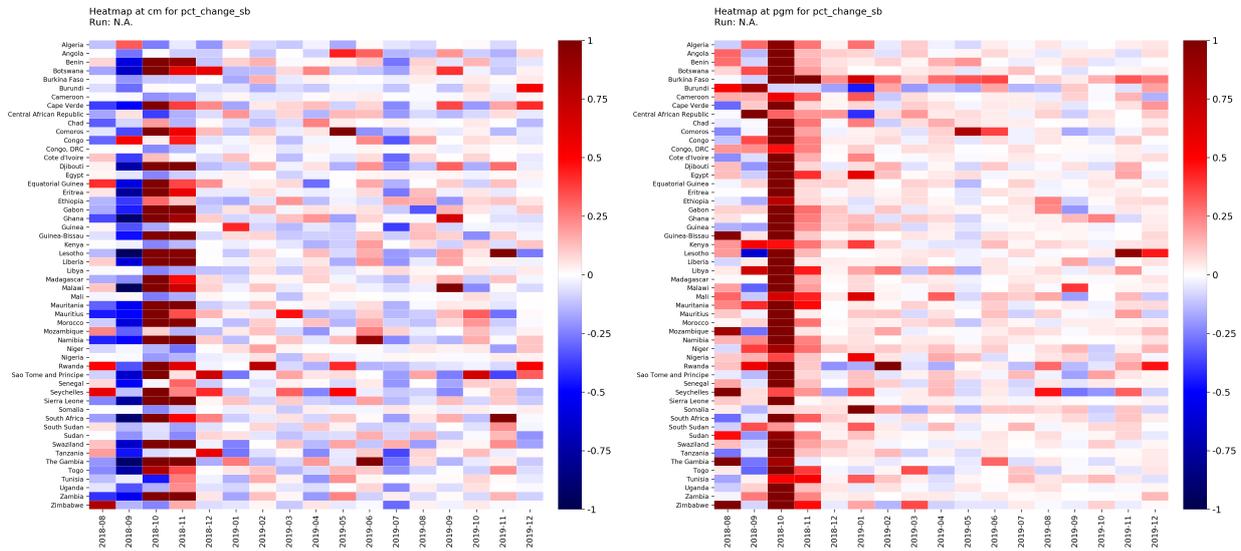

Figure E-11 shows the percentage change in predicted probability of a state-based conflict event at $s = 1$, for each of our forecasts compared to the prior month at the *cm* level (left) and *pgm* level (right), in heat map form. At the *pgm* level, the predicted probability is the proportion of PRIO-GRID cells predicted to have at least one conflict event per country.

REFERENCES

Hegre, Håvard, Marie Allansson, Matthias Basedau, Mike Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina Högbladh, Remco Jansen, Naima Mouhleb, Sayeed Auwn Muhammad, Desirée Nilsson, Håvard Mokleiv Nygård, Gudlaug Olafsdottir, Kristina Petrova, David Randahl, Espen Geelmuyden Rød, Gerald Schneider, Nina von Uexkull, and Jonas Vestby (2019). "ViEWS: A political Violence Early Warning System". In: *Journal of Peace Research* 56.2, pp. 155–174. URL: `https://doi.org/10.1177/0022343319823860`.

Hegre, Håvard, Mihai Croicu, Kristine Eck, and Stina Högbladh (2018). *UCDP-monthly. Monthly updated organized violence data in event and aggregated forms*. Typescript, Uppsala University. `http://www.pcr.uu.se/research/views/`.

Human Rights Watch (2019). *We Found Their Bodies Later That Day*.

Pettersson, Thérése, Stina Högbladh, and Magnus Öberg (2019). "Organized violence, 1989–2018 and peace agreements". In: *Journal of Peace Research* 56.4, pp. 589–603. DOI: `10.1177/0022343319856046`. URL: `https://doi.org/10.1177/0022343319856046`.

Uppsala Conflict Data Program [UCDP] (2019). URL: `https://ucdpged.uu.se/#filter/Geoposition/Burkina20Faso/2018-07-01/2019-09-24`.