Global Food Security 29 (2021) 100510

Contents lists available at ScienceDirect

Global Food Security

journal homepage: www.elsevier.com/locate/gfs

Validating Famine Early Warning Systems Network projections of food security in Africa, 2009–2020

David Backer[a,*], Trey Billing[b]

[a] Center for International Development and Conflict Management, Department of Government & Politics, University of Maryland, USA
[b] Mershon Center for International Security Studies, Ohio State University, USA

ARTICLE INFO

Keywords:
Food security
Early warning
Humanitarian response
Climate
Conflict

ABSTRACT

The Famine Early Warning Systems Network (FEWS NET) operates to mitigate harms associated with food insecurity. Many stakeholders depend on this resource to guide monitoring, planning, interventions, and resource allocations. These activities' effectiveness hinges on the credibility of FEWS NET projections. Published statistical evaluations are rare and narrow in geographic scope. Our extended analysis validates projections for 25 African countries from 2009-2020. Accuracy is 84 percent overall, but drops sharply with ascending food insecurity, biasing toward over-projection. Variation in humanitarian responses, climate, and conflict appear connected to the patterns. The study illuminates FEWS NET's performance in anticipating food insecurity amid fragile conditions and motivates recommendations for improvements through ongoing validation, deeper scrutiny of factors affecting reliability, increased transparency, and informed usage.

1. Introduction

In 2018, the share of people worldwide who were moderately or severely food insecure exceeded 26 percent (FAO 2019). A product of social, economic, and environmental factors (Verdin et al., 2005), food insecurity disproportionately affects people in certain contexts, especially many countries in Africa, where nearly 53 percent of the population is moderately or severely food insecure (FAO 2019). Stakeholders across the international community, down to a local level, depend on monitoring and forecasting tools that assist efforts to mitigate harms associated with food insecurity. The Famine Early Warning Systems Network (FEWS NET) has been a leader in the field since being established by the United States Agency for International Development (USAID) in 1985. FEWS NET produces regular Outlook Reports with current situation assessments and future projections of food security across much of Africa and select countries in Central America, the Caribbean, and Central Asia. These reports are widely used by humanitarian actors to anticipate emergent food security crises and to direct interventions (Ross et al., 2009).

Evaluation of FEWS NET projections is indispensable. Reliable information about food security is essential for the appropriate design and implementation of responses. Knowing about the validity of the projections, as well as appreciating conditions under which validity varies, enables sensible, productive use of the resource. If evaluations reveal issues with the projections, improvements are warranted. Insights from evaluations can guide the evolution of the process FEWS NET employs to generate projections, thereby bolstering their credibility and utility.

Just two published studies present statistical evaluations of FEWS NET projections. Choularton and Krishnamurthy (2019) show that FEWS NET is generally accurate at projecting food security in Ethiopia, though deviations are exacerbated by climate events like El Niño and greatest in the most food-insecure regions. Krishnamurthy, Choularton, and Kareiva (2020) broaden analysis to multiple countries in the Greater Horn of Africa, concluding that armed conflict may also affect deviations from projections of food security, though less so than climatic conditions. These studies represent valuable progress in validation, highlighting achievements and cautioning about shortcomings, which hint at paths of improvement. A limitation is the relatively narrow geographic scope of the analyses, when FEWS NET commonly reports on a diversity of 30-40 countries spanning multiple regions of Africa and the world. No other publicly available research has evaluated the validity of the projections on a larger scale. FEWS NET has conducted its own validation, but decided against releasing the results, instead favoring independent evaluation – according to our communications with members of the leadership. Other examinations of FEWS NET delve into the process by which projections are generated and the underlying data inputs (e.g.,

* Corresponding author.
E-mail addresses: dbacker@umd.edu (D. Backer), billing.22@osu.edu (T. Billing).

https://doi.org/10.1016/j.gfs.2021.100510
Received 10 September 2020; Received in revised form 16 January 2021; Accepted 4 February 2021
Available online 20 March 2021

Brown 2008; Brown and Brickley 2012), without conducting statistical validation.

Our study builds on prior evaluations and fills a gap by analyzing the validity of historical FEWS NET projections for 25 countries (including Yemen) categorized in the East, West, and Southern Africa regions between July 2009 and June 2020. These countries and regions comprise the majority of FEWS NET's coverage worldwide, while the time period reflects the extent of data released as of when this article was written. We conduct separate sets of analyses after converting the data to a granular spatial grid-cell format, with and without weighting by population, as well as at the level of livelihood zones defined by FEWS NET. Data on climatic and conflict conditions are incorporated into the analyses to consider major factors regularly linked to variation in food security that may affect validity. The analysis compares FEWS NET's medium-term projection from a given report cycle to the current situation assessment in the next report. We adopt this approach in lieu of comparing projections against ground-truth data on observable outcomes, an alternative we view as preferable in principle, but deemed infeasible to implement at scale.

The main results from unweighted analysis at a grid-cell level reveal that FEWS NET projections are accurate in around 84 percent of cases. At the lowest level of food insecurity, over 93 percent of projections are accurate. The degree of accuracy declines substantially with ascending levels of food insecurity, exhibiting bias toward over-projection. Further analysis connects over-projections with humanitarian responses during intervals between reporting cycles. Even taking those responses into account, over-projection at more serious levels of food insecurity persists to an extent. To explore potential reasons, we examine relationships to variation in climatic conditions (temperature, precipitation, and vegetation) and the frequency of violent conflict events. The evidence suggests that unanticipated shocks reflected in these factors may hinder the accuracy of FEWS NET projections, running up against constraints of what can be known or realistically foreseen about key drivers of food security.

This study deepens the understanding of FEWS NET's performance in tackling the difficult task of anticipating food insecurity amid settings susceptible to volatility in risks. We conclude with recommendations to enhance the quality of the resource and its usefulness to stakeholders through ongoing validation, consideration of factors affecting reliability, transparency about results of evaluations, and promoting informed utilization.

2. Methods

2.1. Background

From July 2009 through 2015, FEWS NET released Outlook Reports four times per year, in January, April, July, and October. Since 2016, reports have usually been released three times per year, in February, June and October. The exception is December 2018, when reports were released only for a sub-set of regions and countries. A centerpiece of each report is three indices of food security:

- ***CS score***: a current situation assessment as of the month when a report was released.
- ***ML1 score***: a near-term projection for the one month (in December 2018 reports), two months (in reports from July 2009 through October 2015) or three months (in the normal cycle of reports since February 2016) immediately after the release month.
- ***ML2 score***: a medium-term projection for a period 3-5 months (in reports from July 2009 through October 2015), 3-6 months (in December 2018 reports), or 4-7 months (in the normal cycle of reports since February 2016) subsequent to the release month.

Throughout the timeframe that our analysis covers, FEWS NET assigned index scores on a 5-level scale, numbered in order of ascending food insecurity (see Table 1). Fig. A1 in Section A1.1 of the Appendix summarizes the structure of the data, which varies according to the frequency of reporting and periodicity of projections.

Table 1
FEWS NET index of food insecurity. The index scale has been compatible with the Integrated Phase Classification (IPC) since April 2011. Previously, FEWS NET employed a distinctive approach and labelled the levels slightly different from the IPC, but they were largely analogous.

Level	FEWS NET Label (July 2009–January 2011)	IPC Label (April 2011 – present)
5	Famine	Catastrophe/Famine
4	Extremely food insecure	Emergency
3	Highly food insecure	Crisis
2	Moderately food insecure	Stressed
1	No acute food insecurity (2009: generally food secure)	Minimal/none

2.2. Framework of validation

Our evaluation compares ML2 projections from a given report to CS assessments from the next report. The logic is that a projection represents a "prediction" and a corresponding assessment represents an "outcome" for an overlapping time period. For example, ML2 projections in the February 2020 report covered the period from June–September 2020, while the June 2020 report (the latest included in our analysis) provided updated CS assessments for that same month. Similar temporal overlaps between the two measures occur at multiple points in each year from 2009 onwards.

A variant on the approach would be to compare ML2 scores in a report to ML1 scores in the next report. In our communications with leadership, we learned that FEWS NET favors such a comparison, which was employed in its unreleased validation analysis, as a means to maximize temporal overlap between measures. A drawback is that both measures reflect projections into the future, which is not optimal for evaluating predictive performance. Our approach avoids this weakness, while yielding substantively equivalent findings (results not shown).

A further consideration is that either of these approaches involves internal validation: one FEWS NET measure is compared against another FEWS NET measure. An implicit risk is a data-generating process that is prone to attenuate differences between the measures. One possible reason could be insufficient independence across reporting cycles that artificially amplifies serial autocorrelation. Experts also raise concerns about manipulation – especially under political influence – of both information inputs and analytical outputs of food security early warning systems (e.g., Maxwell and Hailey 2020). Investigating the integrity of FEWS NET's process is outside the scope of our study, which concentrates on analysis of the data as reported. Nevertheless, if either of these dynamics is in play, our approach intrinsically skews against finding inaccuracies. Therefore, the results should overstate the degree of accuracy, at worst, elevating the significance of any inaccuracies we do find. A countervailing dimension, however, is that the integer scale of the index raises the chances of finding inaccuracies (since scores must match exactly to be considered accurate) and a greater extent of bias.

Another alternative is to compare FEWS NET projections to empirical data on outcomes. This sort of external validation, using ground truth as a benchmark, is a traditional gold-standard for evaluating the accuracy of predictions. An inherent challenge is that FEWS NET gauges food security, which is a subjective construct that cannot be directly measured. Other indices of food security are calculated based on sets of indicators, and related measures (e.g., childhood malnutrition or caloric intake) do exist. Differences in how they are conceptualized and computed undermine comparability with FEWS NET projections. Also, comprehensive historical data with sufficient spatio-temporal coverage and granularity for potential benchmarks are not readily available. Since an extensive scope is central to our aims, we opted against selective

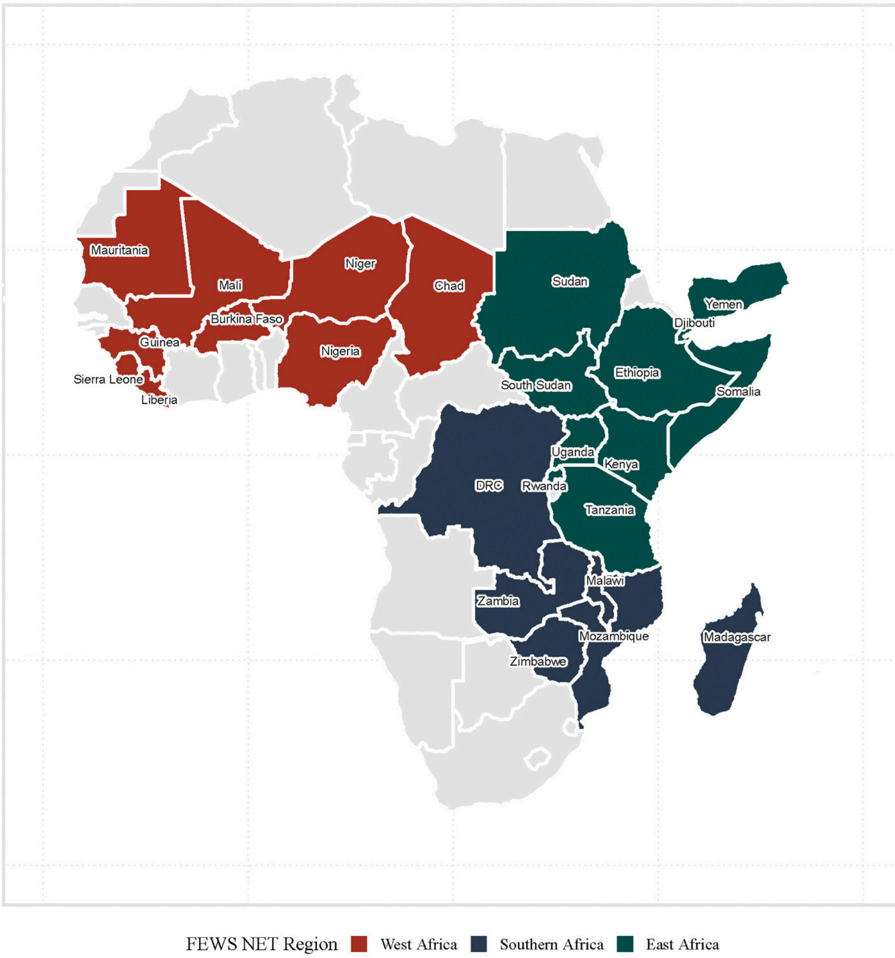

Fig. 1. Geographic scope of validation. FEWS NET released multiple Outlook Reports about food security from 2009-2020 in the 25 countries highlighted on the map. FEWS NET groups these countries by region. *East Africa* (shaded in green) includes Djibouti, Ethiopia, Kenya, Rwanda, Somalia, South Sudan, Sudan, Tanzania, Uganda, and Yemen. *Southern Africa* (shaded in blue) includes the Democratic Republic of Congo, Madagascar, Malawi, Mozambique, Zambia, and Zimbabwe. *West Africa* (shaded in orange) includes Burkina Faso, Chad, Guinea, Liberia, Mali, Mauritania, Niger, Nigeria, and Sierra Leone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

external validation with a limited scope, while endorsing this avenue as worthwhile for follow-up research.

2.3. Data processing

FEWS NET's online Data Center archives the Outlook Reports and associated geographic information system (GIS) shapefiles that reflect assessments and projections of food security dating back to July 2009 (see https://fews.net/data). From this source, we obtained all the shapefiles available as of September 1, 2020 associated with reports released for the East, West, and Southern Africa regions (see Fig. 1), which cover up through June 2020. FEWS NET includes Yemen as part of the East Africa region.

These data are used to conduct three sets of analyses. Our main set of analyses relies on converting the data to the uniform, static format of $0.5^\circ \times 0.5^\circ$ grid-cells. The primary advantages are comparability among these spatial units, both in size (roughly $55\,km^2$) and especially over time. The Appendix goes into depth with an explanation of the rationale for the conversion, a description of the mechanics, and an illustrative example (see Fig. A2 in Section A1.2). Two additional sets of analyses serve as checks of the sensitivity of the results to the choice of design. One set of analyses is conducted at the level of livelihood zones. A justification is that these zones are the actual units for which FEWS NET gauges food security. Yet the zones vary both in size and over time, which affects comparability. Moreover, we encountered complications in disaggregating shapefiles of assessments and projections to the level of these zones, as well as in matching cases over time – prerequisites of the analysis. Another set of analyses weights the grid-cells by population. A primary motivation is that the density of population can vary significantly across areas of countries. FEWS NET and stakeholders have more at stake with projections made for densely populated areas than those made for sparsely populated areas, since density affects expected caseloads of people who suffer impacts of food insecurity. A limitation, however, is that population data with necessary spatio-temporal granularity are lacking. Presenting the three sets of analyses, with countervailing pros and cons, ensures a more robust validation.

2.4. Validation metrics

We employ three metrics when validating FEWS NET projections. First, ***accuracy*** is the share of grid-cells for which a given ML2 projection matched the next CS assessment. This metric is common in evaluations of the performance of predictive models involving classification among discrete outcomes. Projecting ordinal levels of food security is an example of classification. Second, we follow Choularton and Krishnamurthy (2019) and calculate ***bias*** as the mean difference between assessments and projections across grid-cell cases. Third, ***absolute deviation*** is the mean absolute difference between assessments and projections across cases. This metric conveys the extent of deviations, regardless of sign. By contrast, positive and negative deviations can cancel out one another in the calculation of bias, leaving a misleading overall impression of close correspondence between assessments and projections. For the purposes of our analysis, therefore, accuracy captures how often projections are exactly right, bias captures whether projections lean toward overshooting or undershooting, and absolute deviation captures whether projections tend to be off by a lot or a little.

Table 2
Validation metrics for full sample. The levels are IPC compatible since April 2011. Definitions and interpretations of accuracy, bias, and absolute deviation are elaborated in the text. The sample covers all reports produced by FEWS NET related to 25 countries in Africa (see Fig. 1 for list) from July 2009–June 2020.

Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	29.21%	-0.90	0.90	0.08%
4	41.23%	-0.66	0.67	1.92%
3	65.84%	-0.26	0.37	11.18%
2	74.38%	-0.03	0.26	27.24%
1	92.65%	0.08	0.08	59.58%
Overall	83.64%	0.00	0.17	100.00%

Our statistical analysis is descriptive in nature, rather than inferential. We characterize patterns in the data, without formally testing propositions about why patterns are expected to be observed.

3. Results

Based on our analysis at the grid-cell level unweighted by population, the overall accuracy of medium-term projections of food security by FEWS NET from July 2009 to June 2020 was nearly 84 percent, as shown in Table 2. Bias was essentially zero, whereas the absolute deviation was ± 0.17 of a level on average. Of note, values of all the metrics vary as a function of the level of projected food security. Nearly 93 percent of grid-cells projected at level 1 were accurately classified. As projections ascend the scale, accuracy drops precipitously. Only 66 percent of grid-cells projected at level 3, a minority projected at level 4, and less than 30 percent projected at level 5 were accurately classified. These cases comprise small shares of the total number, but stand out given the seriousness of food insecurity being projected and the stakes inherent in responses. Meanwhile, bias turns negative and becomes increasingly so, while the degree of absolute deviations also becomes more pronounced, as projections ascend the scale.

The Appendix presents analogous results of analyses conducted at the livelihood zone level [see Table A1 in Section A2.1] and at the grid-cell level weighted by population [see Table A2 in Section A2.2]. Both sensitivity checks yield findings substantively similar to those for the analysis conducted at the grid-cell level without weighting by population. Therefore, we focus in the rest of this section on results from unweighted grid-cell level analyses.

Table 3 cross-tabulates the shares of grid-cells that fall in each assessment level (rows) relative to each projection level (columns). Within this contingency table, the diagonal that runs from the lower left to the upper right reflects accurately classified cells, while over-projections are below the diagonal and under-projections are above the diagonal. Aside from at level 1 (by default), shares of over-projections outweigh shares of under-projections. Another favorable finding is that less than 1 percent of grid-cells deviate by multiple levels from projections. The rare cases are noteworthy because of the repercussions for the efficiency of humanitarian responses, as well as perceptions of FEWS NET's reliability. Projections at level 3 and above will ordinarily prompt alerts, which could appear unjustified if subsequent assessments of food insecurity turn out to be two or even three levels less severe.

Meaningful differences in validity arise across regions (see Fig. A5 and Table A3 in Section A2.3 of the Appendix). Accuracy is comparably high for West Africa (87 percent) and Southern Africa (86 percent), but appreciably lower for East Africa (75 percent). In all three regions, accuracy is 90 percent or above for projections at level 1. Projections at level 3 were accurate for 60–70 percent of grid-cell cases in each region. Accuracy falls to around 40 percent for projections at level 4 in East and West Africa. Such projections were rare in Southern Africa and never accurately classified. Projections at level 5 are limited to East Africa and accurate below 30 percent of the time. Bias for both East Africa and West Africa is close to zero, while bias for Southern Africa reveals a slight tendency toward over-projection. Absolute deviation is more pronounced for East Africa than either West or Southern Africa. Notable differences in results are also evident across countries, including among those within the same region (see Fig. A6 in **Section A2.4** of the Appendix). Furthermore, results vary markedly at the most granular level,

Table 3
Contingency table for full sample. This cross-tabulation presents the distribution of current situation assessments in a given FEWS NET report (in rows) relative to the corresponding medium-term projections in the previous FEWS NET report (in columns). Each cell within the table displays the share of cases (top value) and number of cases (bottom value in brackets), by the level of the projection. A case corresponds to a grid-cell unit for a given FEWS NET cycle. Darker shades of blue indicate higher shares of cases. The levels are IPC compatible since April 2011. The sample covers 25 countries tracked by FEWS NET in Africa (see Fig. 1 for list) from July 2009–June 2020.

Level of Next FEWS NET Current Situation Assessment					
5	0.00% [N=0]	0.00% [N=0]	0.12% [N=45]	0.19% [N=12]	29.21% [N=78]
4	0.00% [N=0]	0.17% [N=148]	4.91% [N=1805]	41.23% [N=2612]	52.06% [N=139]
3	0.45% [N=915]	11.19% [N=10028]	65.84% [N=24222]	51.22% [N=3245]	18.73% [N=50]
2	6.88% [N=13487]	74.38% [N=66676]	26.92% [N=9905]	6.93% [N=439]	0.00% [N=0]
1	92.65% [N=181667]	14.27% [N=12795]	2.21% [N=812]	0.43% [N=27]	0.00% [N=0]
	[N=196069]	[N=89647]	[N=36789]	[N=6335]	[N=267]
Level of FEWS NET Medium-Term Projection	1	2	3	4	5

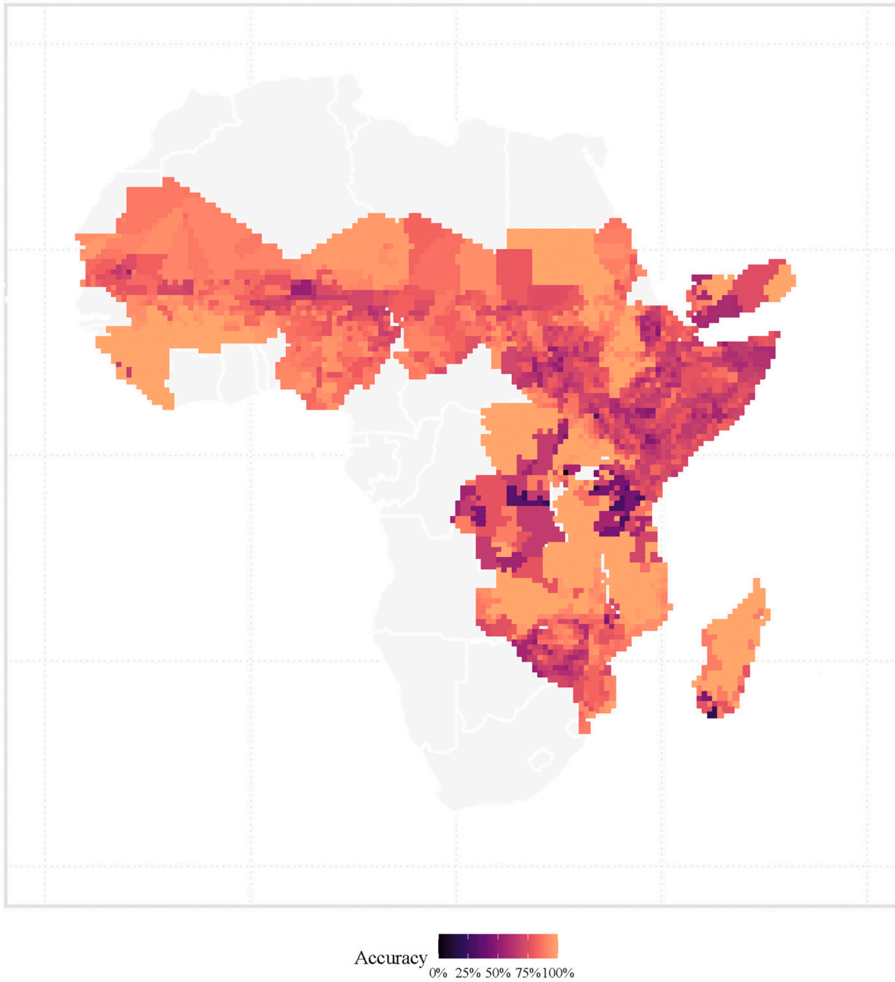

Fig. 2. Overall accuracy in FEWS NET projections by grid-cell. Accuracy is defined as the share of grid-cell cases in which the current situation assessment of food security from a given report matches the corresponding medium-term projection from the previous report. The calculations cover the entire sample of cases, spanning 25 countries tracked by FEWS NET in Africa from July 2009–June 2020.

as is displayed in maps of the overall mean accuracy (Fig. 2), bias (Fig. 3), and absolute deviation (Fig. 4) by grid-cell across the entire time period covered in the analysis.

In addition, results vary to an extent over time (see Fig. A7 in **Section A2.5** of the Appendix). Accuracy, bias, and absolute deviation for projections at level 1 were mostly stable, though trending slightly in a worse direction over recent years. Accuracy of projections at level 2 peaked around 2014 and has been trending downward since, accompanied by steady increases in bias and absolute deviation. The metrics for projections at level 3 have gradually trended in a favorable direction. The trends for projections at level 4, while not monotonic, reveal dramatic shifts toward lower accuracy, greater negative bias, and increased absolute deviation over the long run (See also Fig. A8 in **Section A2.5** of the Appendix, which presents a time-lapse animation of the trend in bias by grid-cell from report to report.).

3.1. Relating projection performance to humanitarian assistance

Since April 2012, FEWS NET data indicate CS assessments that "would likely be at least one phase worse without current or programmed humanitarian assistance." In the context of our analysis, assistance is flagged for 5.45% of the grid-cells across the regions, ranging from 8.41% in East Africa to just 2.19% in Southern Africa and 1.71% in West Africa.

We assume that this assistance is an exogenous factor, which was unanticipated when a prior ML2 projection was made. By definition, the assistance reduces the level of the subsequent CS assessment against which the projection is compared. Consequently, our intuition is that flagged cases should tilt in the direction of lower accuracy, more negative bias, and greater absolute deviation for projections aside from those at level 1 (where effects should be reversed), relative to grid-cells not flagged with assistance. The impact may be magnified for projections at higher levels of food insecurity, in so far as these projections correlate with heightened humanitarian responses that improve assessments by more than a single level.

Table 4 presents the results of the validation metrics for grid-cells with (top panel) and without (bottom panel) humanitarian assistance flags. Table 5 provides a contingency table conditional on the flag (See also Table A3 in Section A2.3 of the Appendix for the corresponding contingency table by region.).

For projections at level 1, accuracy without humanitarian assistance exceeded accuracy with assistance – a surprising result. An association between assistance and greater accuracy might be expected, since cases that would otherwise be assessed at level 2 or above, in the absence of assistance, stand a better chance of being at level 1 due to the assistance. An explanation is the grid-cells flagged with assistance after being projected at level 1 experienced shocks that unexpectedly worsened food security. The assistance mitigates the situation, but not always

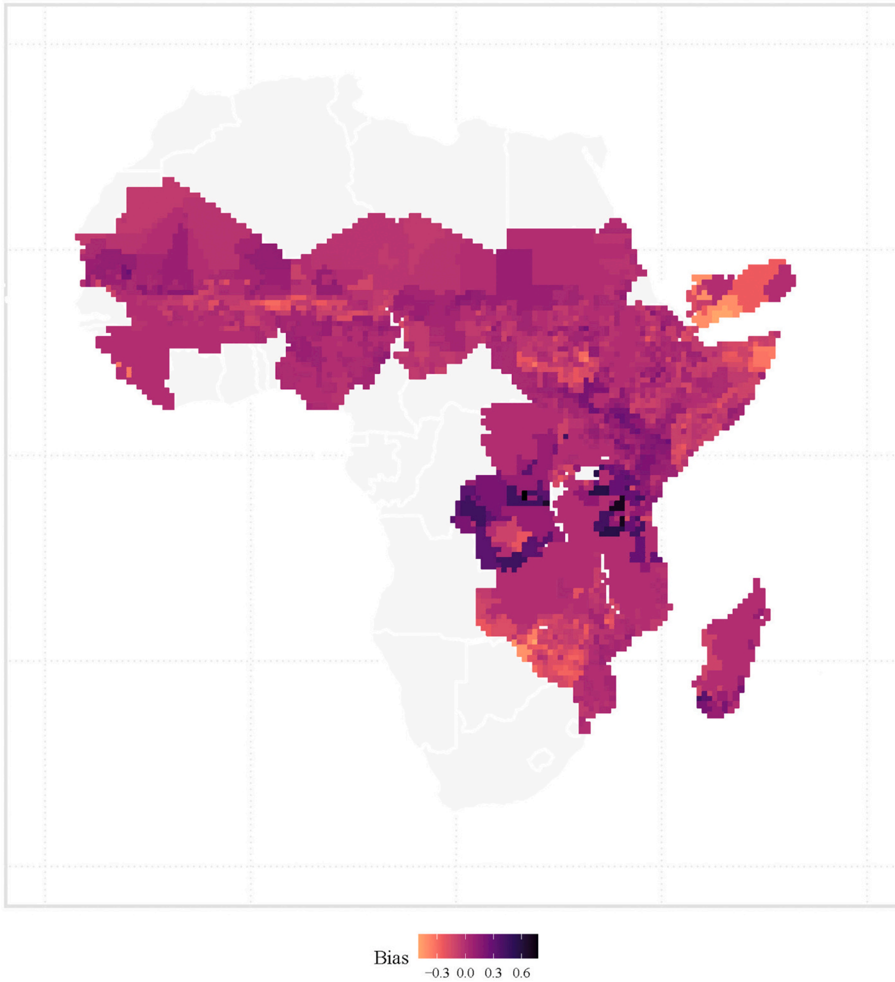

Fig. 3. Overall bias in FEWS NET projections by grid-cell. Bias reflects the current situation assessment of food security from a given report minus the corresponding medium-term projection from the previous report. The calculations cover the entire sample of cases, spanning 25 countries tracked by FEWS NET in Africa from July 2009–June 2020.

completely, with many cases ending up at level 2.

For projections at level 2, accuracy drops off, bias is negative, and the absolute deviation is higher for the grid-cells without humanitarian assistance, whereas accuracy actually increases, bias is nearly 0, and the absolute deviation is lower for the grid-cells with humanitarian assistance. These results hint at the effectiveness of assistance in settings projected to be "Stressed" by food insecurity. The high accuracy rate hides the favorable impact of assistance, without which accuracy would have been lower. A possible interpretation is that FEWS NET's projections helped to precipitate successful humanitarian responses.

For projections at level 3, accuracy is 35 percentage points lower among grid-cells flagged with humanitarian assistance, relative to those not flagged. The nature of bias for flagged cases implies that in the absence of assistance, FEWS NET would have under-projected food insecurity on average (given assessments at least one level worse). Instead, the net effect of the assistance is to indicate over-projection. These results are revealing about the middle range of cases for which humanitarian responses appear to be hard to calibrate. Over 56 percent of these cases turn out better than projected and another 42 percent wind up no worse than projected – as a result of assistance. These outcomes are arguably worthwhile, even if some inefficiencies exist.

For projections at level 4 and 5, the tendency toward humanitarian assistance resulting in better-than-expected outcomes is accentuated. The accuracy of less than 1 percent and values for bias and absolute deviation indicate that no grid-cells turn out worse than expected and very few as bad as expected. Almost 86 percent of grid-cells projected to be level 4s, then flagged with humanitarian assistance, are subsequently assessed at level 3 and another 14 percent at level 2. These results compare to 52 percent of projections at level 4 that are accurately classified in the absence of humanitarian assistance. Those cases point to the limits of FEWS NET projections in prompting responses that alleviate food insecurity. Projections about risks of an "Emergency" level of food insecurity are more often right than wrong. Still, fewer grid-cells facing such projections benefited from humanitarian assistance that reduced food insecurity than grid-cells where either a response was ineffective or did not occur.

3.2. *Relating projection performance to climatic conditions*

Next, we explore whether the validity of FEWS NET projections is related to deviations from normal climatic patterns. The premise is that the process FEWS NET employs may not account adequately for climate variation. According to a survey by Brown and Brickley (2012), FEWS NET analysts consult data on rainfall when generating reports over 80 percent of the time and data on vegetation almost 30 percent of the time. The rates below 100 percent suggest that analysts overlook climatic conditions – extreme or otherwise – at least some of the time. Analysts must also make judgements about likely effects of future climate

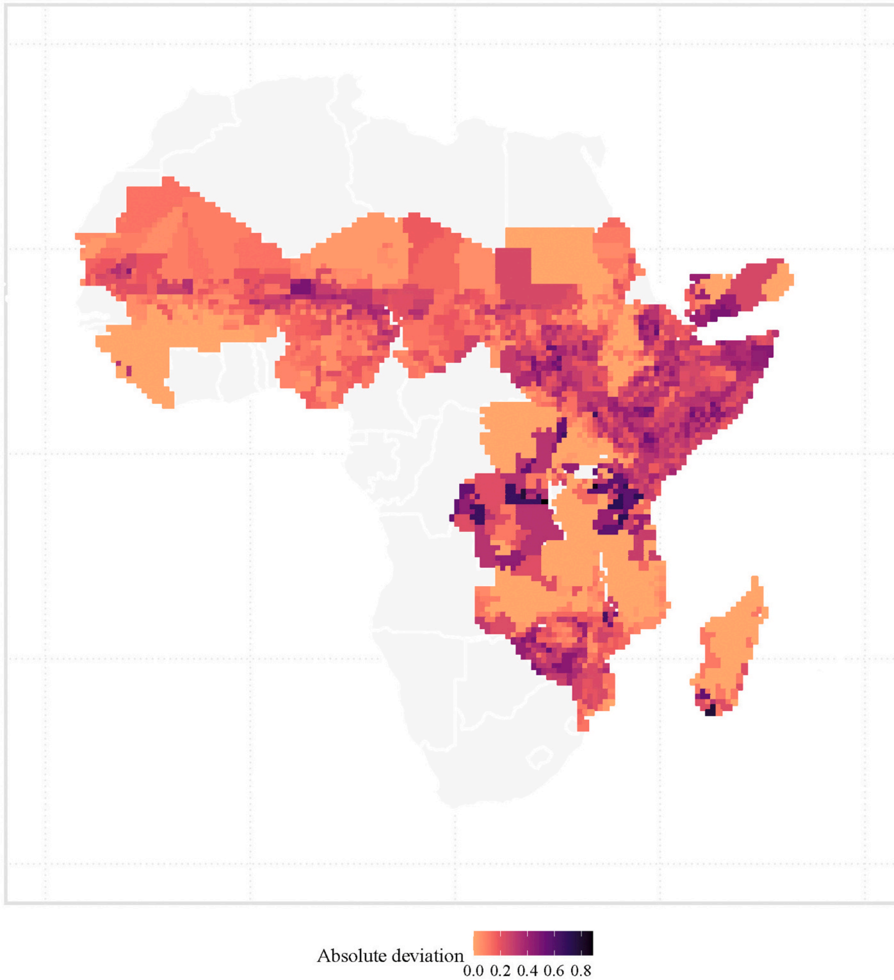

Fig. 4. Overall absolute deviation in FEWS NET projections by grid-cell. Absolute deviation reflects the absolute value of the difference between the current situation assessment in a given report and the corresponding medium-term projection in the previous report. The calculations cover the entire sample of cases, spanning 25 countries tracked by FEWS NET in Africa from July 2009–June 2020.

Table 4
Validation metrics conditional on humanitarian assistance. The levels of food security are IPC compatible. Humanitarian assistance reflects the cases as flagged by FEWS NET. Definitions and interpretations of accuracy, bias, and absolute deviation are elaborated in the text. The sample covers 25 countries in Africa tracked by FEWS NET from April 2012–June 2020.

With Humanitarian Assistance Flag				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	0.00%	-2.00	2.00	0.04%
4	0.25%	-1.14	1.14	9.03%
3	42.27%	-0.59	0.59	26.38%
2	89.17%	-0.02	0.11	52.87%
1	70.15%	0.31	0.31	11.68%
Overall	66.51%	-0.23	0.35	100.00%
Without Humanitarian Assistance Flag				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	0.00%	-1.00	1.00	0.00%
4	51.61%	-0.53	0.53	1.03%
3	77.48%	-0.17	0.25	9.21%
2	76.18%	-0.01	0.24	23.12%
1	93.35%	0.07	0.07	66.65%
Overall	87.49%	0.03	0.13	100.00%

conditions, without necessarily having reliable forecasts several months ahead (see FEWS NET 2018b, which outlines how analysts are instructed to use rainfall data). Judgments can be off for multiple reasons. Even when analysts consider climate indicators, accuracy of projections may be affected by unexpected shocks that cause actual climate patterns to depart from expectations. Analysts may overstate or understate the impact of climate. The direction of resulting bias will depend on specific climate factors and relationships to food insecurity that the analyst assumes.

For each projection, we calculate the difference (expressed in units of standard deviations) between climatic conditions during the month of the next report relative to the mean for that same month over the last five years – all at the grid-cell level. Our assumption is that the long-run historical mean affords a reasonable approximation of the conditions expected over the time period covered by the projection, while conditions contemporaneous with the next report are likely to influence the CS assessment. We use publicly available, high-resolution, monthly data to analyze three standard indicators of climate: *temperature* ($CHIRTS_{max}$; Funk et al., 2019), *precipitation* (CHIRPS; Funk et al., 2015), and *vegetation* (NDVI; Vermote, 2019). The analyses also account for humanitarian assistance flags.

As displayed in the top panel of Fig. 5, over-projection at higher levels of food insecurity is a pattern across the range of temperatures for cases flagged with humanitarian assistance. Otherwise, hotter than normal conditions (>1 standard deviation) are largely associated with

Table 5
Contingency table conditional on humanitarian assistance. These cross-tabulations present the distribution of current situation assessments in a given FEWS NET report (in rows) relative to the corresponding medium-term projections in the previous FEWS NET report (in columns), for cases with and without humanitarian assistance flagged by FEWS NET. Each cell within the table displays the share of cases (top value) and number of cases (bottom value in brackets), by the level of the projection. Total cases by the level of projection are also listed at the bottom of each column. A case corresponds to a grid-cell unit for a given FEWS NET cycle. Darker shades of blue indicate higher shares of cases. The levels are IPC compatible. The sample covers 25 countries in Africa tracked by FEWS NET from April 2012–June 2020.

With Humanitarian Assistance Flags

Level of Next FEWS NET Current Situation Assessment	1	2	3	4	5
5	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]
4	0.00% [N=0]	0.01% [N=1]	0.09% [N=4]	0.25% [N=4]	0.00% [N=0]
3	0.64% [N=13]	4.47% [N=413]	42.27% [N=1948]	85.54% [N=1349]	100.00% [N=7]
2	29.22% [N=596]	89.17% [N=8235]	56.21% [N=2590]	14.20% [N=224]	0.00% [N=0]
1	70.15% [N=1431]	6.35% [N=586]	1.43% [N=66]	0.00% [N=0]	0.00% [N=0]
	[N=2040]	[N=9235]	[N=4608]	[N=1577]	[N=7]

Level of FEWS NET Medium-Term Projection

Without Humanitarian Assistance Flags

Level of Next FEWS NET Current Situation Assessment	1	2	3	4	5
5	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	0.13% [N=3]	0.00% [N=0]
4	0.00% [N=0]	0.14% [N=70]	3.70% [N=752]	51.61% [N=1169]	100.00% [N=4]
3	0.54% [N=788]	11.46% [N=5846]	77.48% [N=15746]	44.37% [N=1005]	0.00% [N=0]
2	6.11% [N=8992]	76.18% [N=38878]	16.50% [N=3353]	3.00% [N=68]	0.00% [N=0]
1	93.35% [N=137324]	12.22% [N=6238]	2.32% [N=472]	0.88% [N=20]	0.00% [N=0]
	[N=147104]	[N=51032]	[N=20323]	[N=2265]	[N=4]

Level of FEWS NET Medium-Term Projection

unbiased projections, whereas cooler than normal conditions are associated with over-projection at higher levels of food insecurity. The middle panel of Fig. 5 reveals over-projection at higher levels of food insecurity throughout the range of precipitation when humanitarian assistance is flagged. Otherwise, drier than normal conditions (<-1 standard deviation) are more often associated with unbiased projections of food insecurity, whereas wetter conditions are associated with over-projection at higher levels of food insecurity. Results for deviations in vegetation, shown in the bottom panel of Fig. 5, mirror those for precipitation, which makes sense given the relationship between the factors. Analogous results of analyses conducted at the grid-cell level weighted by population are substantively similar [see Fig. A3 in Section A2.2 of the Appendix].

The upshot is favorable climatic shocks appear to be part of an explanation for why FEWS NET over-projects severe food insecurity with some regularity. The results offer little evidence of a converse pattern linking unfavorable shocks with under-projections of levels of food insecurity, though this finding may be due to intervening effects of humanitarian assistance.

3.3. Relating projection performance to conflict conditions

The final step in our analysis is to examine whether the accuracy of FEWS NET's projections is related to patterns of conflict activity. When generating measures of food security, FEWS NET analysts do consider conflict as a factor (FEWS NET, 2018a). Yet conflict can unfold unexpectedly and judgements about this factor's influence based on available data may not be satisfactory.

We focus on the potential impact of violent conflict events. For the purpose, we integrate multiple sources that collectively capture diverse

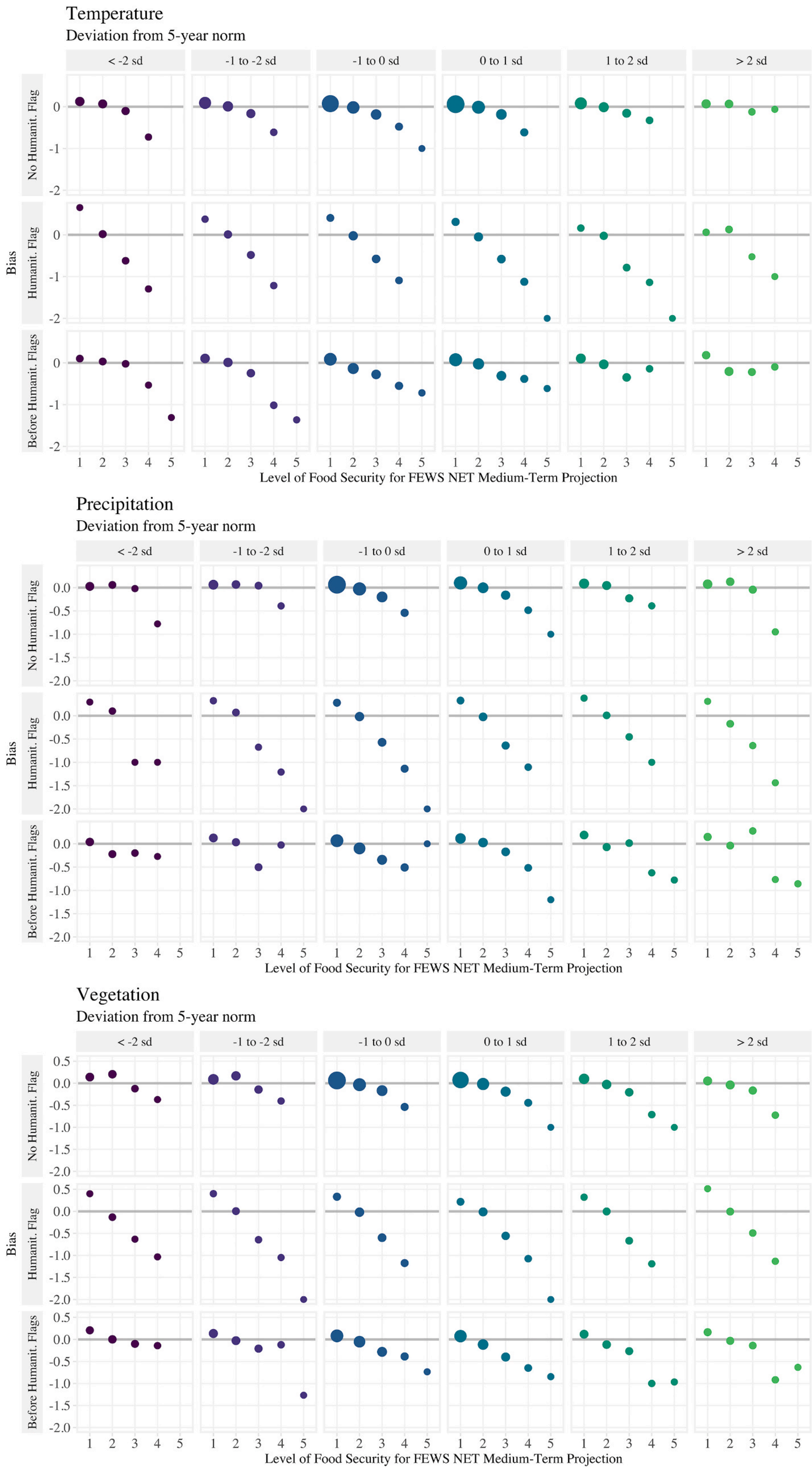

Fig. 5. Bias in FEWS NET projections by deviations in climatic conditions and conditional on humanitarian assistance. Standard deviations (sd) reflect the observed value of a climatic indicator for the month of the next FEWS NET report after a given projection, relative to the observed values of the indicator for the same month over the previous five years. The size of bubbles is proportionate to the share of grid-cell cases. The sample includes 25 countries tracked by FEWS NET in Africa. The top and middle sets of graphs in each panel cover April 2012–December 2018, whereas the bottom set of graphs in each panel covers July 2009–January 2012.

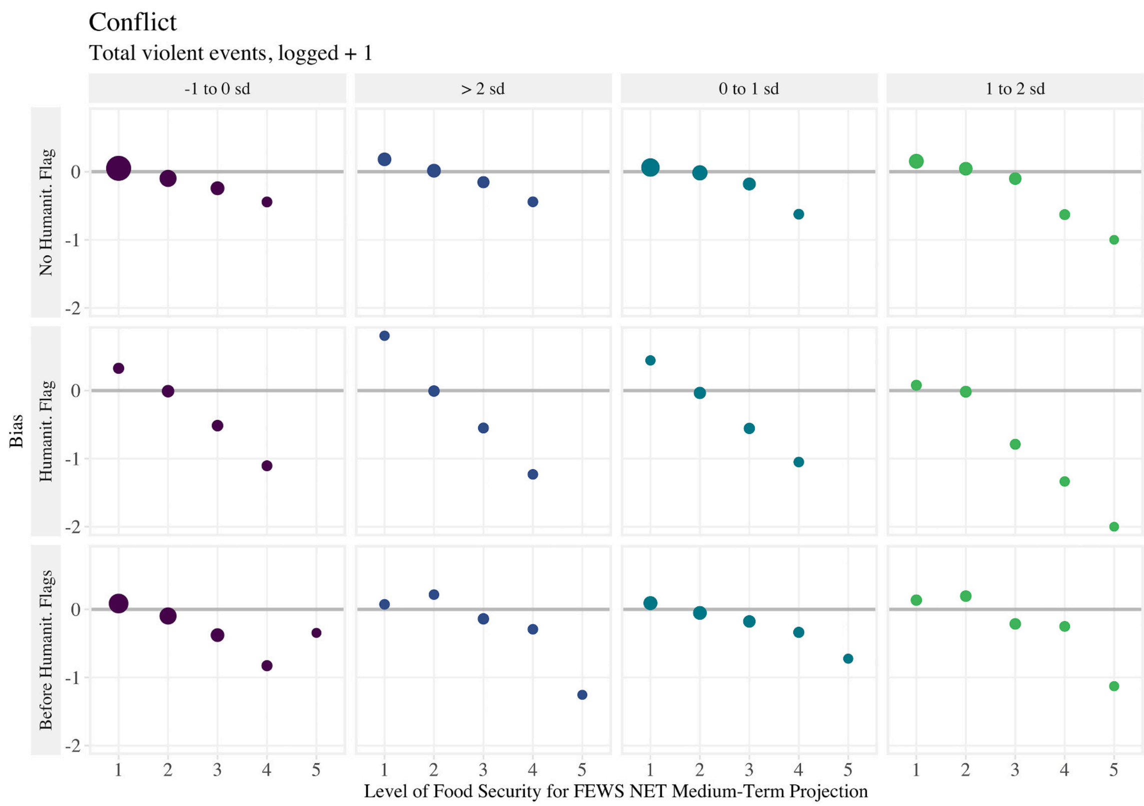

Fig. 6. Bias by frequency of violent conflict events conditional on humanitarian assistance. Standard deviations (sd) reflect the observed value of violent conflict events for the month of the next FEWS NET report after a given projection, relative to the observed values for the same month over the previous five years. The size of bubbles is proportionate to the share of grid-cell cases. The sample includes 25 countries tracked by FEWS NET in Africa. The top and middle sets of graphs cover April 2012–December 2018, whereas the bottom set covers July 2009–January 2012.

types of events: the Armed Conflict Location and Event Data (Raleigh et al., 2010), the Global Terrorism Database (START, 2019), and the Uppsala Conflict Data Program Georeferenced Dataset (Sundberg and Melander, 2013). The integration employs the Merging Event Data by Location, Time, and Type (MELTT) software package (Donnay et al., 2019), after we first subset each dataset to extract only events resulting in fatalities. All the datasets georeference and time stamp events, which enables a spatio-temporal merge with the FEWS NET data. As an indicator of violent conflict, we use the frequency of events during the month of the report subsequent to a given projection. The analysis is confined to 2009–2018, based on current availability of conflict event data from all three sources.

As seen in Fig. 6, higher frequencies of violent conflict events in the absence of humanitarian assistance are associated with more consistent patterns of unbiased projections of all levels of food insecurity. In grid-cells flagged with humanitarian assistance, violent conflict events are associated with over-projection at higher levels of food insecurity, but also under-projection at the lowest level. The latter result is noteworthy, implying that conflict activity in places where food insecurity was of least concern tended to worsen outcomes unexpectedly, likely precipitating a humanitarian response, which was insufficiently effective. Analogous results of analyses related to conflict conditions conducted at the grid-cell level weighted by population are substantively similar [see Fig. A4 in **Section A2.2** of the Appendix].

4. Discussion

Until recently, FEWS NET projections of food security had not been subjected to statistical evaluations, reported publicly, on which stakeholders can rely for orientation about the credibility of this resource. Two studies published since 2019 made vital gains by investigating the internal validity of projections in select countries when compared to FEWS NET's own subsequent current situation assessments. Our analysis follows in those footsteps, accomplishing a consequential advance by undertaking the first large-scale validation of historical projections, extending the scope to encompass all 25 countries FEWS NET tracked in Africa (including Yemen) over more than a decade. We determine that FEWS NET's medium-term projections are accurate most of the time, again judged against subsequent current situation assessments. The overall results hide a crucial nuance: accuracy is exceptional at the lowest projected levels of food insecurity, but drops off substantially with ascending levels, tilting toward over-projection. These findings are consistent with what Choularton and Krishnamurthy (2019) detected in their initial study of Ethiopia. Our analysis associates the tendency with the impact of humanitarian assistance, which projections of food insecurity by FEWS NET may help to spur. Further results suggest that unanticipated climate and conflict shocks contribute to diminishing the accuracy of projections. Here too, we corroborate findings of the foundational studies. Like Choularton and Krishnamurthy (2019), we see signs that fluctuations in climatic conditions contribute to inaccuracies in projections. Our findings are also consistent with the conclusion of

Krishnamurthy et al. (2020) that climatic factors exert a stronger influence on these deviations than do circumstances of conflict. A primary value-added of our study is generalizing constructive insights from previous research across an expansive array of countries with diverse characteristics in Africa.

5. Conclusion

Early warning has become an essential tool for assisting efforts to forestall crises with devastating consequences, including famines. FEWS NET deserves respect for a lengthy, estimable record of tackling the hard task of anticipating complex food security outcomes in challenging settings around the world. Stakeholders need to have confidence about the accuracy of FEWS NET projections. Methodical evaluation with a critical lens is integral in that regard. Our study, conducted in this spirit, achieves useful contributions by enriching an emergent line of inquiry, highlighting basic facts about the validity of projections, and probing the roles of key drivers of food security. What we encountered and gleaned when undertaking the study leads us to offer several recommendations, directed with respect to fellow researchers, FEWS NET, and stakeholders – especially humanitarian actors.

First, we advocate continued attention to validation of FEWS NET (and other similar early warning initiatives). Our analyses make important headway, but are not sufficient let alone definitive. An obvious extension, to be as comprehensive as possible, is to maximize the geographic scope of validation by covering all the countries FEWS NET tracks around the world. Another design option we endorse is external validation of projections against ground-truth benchmarks of food security or related proxies. Additional investigation of factors that could affect the accuracy of projections is also worthwhile. We prioritized climatic and conflict conditions, which are attributed as prominent causes of food security crises, as well as humanitarian assistance, which is intended to mitigate vulnerabilities. Evaluations could examine factors fundamental to the equation such as food production, market prices, health, behaviors of populations, demographics, institutional environments, and physical geography. The goals would be to pinpoint sources of bias and the lack of improvement in accuracy over time that are overlooked or inadequately captured by FEWS NET, or else built into its system of generating assessments and projections.

Second, we encourage FEWS NET's support of validation. Our firsthand experiences indicate an openness to engagement around evaluation and to introspection about implications of results of analyses, which is promising. Findings of evaluations should inform deliberations about changes to the way projections are generated. We have seen FEWS NET demonstrate a willingness to adapt by integrating new data, expertise, and resources (e.g., model-based forecasts of acute malnutrition prevalence rates) into the existing workflow. To foster validation, FEWS NET could also release the more granular distributions of probabilities that underlie the projections, which would enable replication and provide a better handle on uncertainty. In addition, we urge an overt commitment to transparency about the validity of FEWS NET projections. Results from evaluations should be compiled, made available, and publicized – and treated as required reading alongside the reports.

Third, we advise stakeholders to be knowledgeable and savvy consumers of FEWS NET products. To use the resource adroitly, stakeholders ought to be attuned to the findings about the validity of the projections – and the nature of variation in accuracy, bias, and uncertainty.

Fourth, we propose more work to ascertain the effectiveness of the loop connecting FEWS NET projections to humanitarian assistance. Patterns we identify in the data seem to uphold an argument that the validity of projections is strengthened as a function of the consequences of ensuing assistance, which is no coincidence. Establishing the association conclusively would bolster the claim.

Funding

This article is a product of a project funded by the UK Aid from the United Kingdom Foreign, Commonwealth & Development Office. The work of the authors was supported via a sub-contract from Action Against Hunger to the University of Maryland. The views expressed in the article do not necessarily represent official policies of the UK Government.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors appreciate valuable comments from the editor and reviewers of *Global Food Security* and extensive feedback about earlier versions of the research by members of the Technical Advisory Group and collaborators on the Modeling Early Risk Indicators to Anticipate Malnutrition (MERIAM) project, participants in the May 2019 workshop on "Next-Generation Food Shock Modeling" sponsored by the Aspen Global Change Institute, and staff of the Famine Early Warning Systems Network and the United Nations Office for the Coordination of Humanitarian Affairs.

Appendix

This Appendix provides further details related to the methods (data structure and processing), as well as results of supplementary analyses (by livelihood zone, population-weighted, by region, by county, over time).

A1. Methods

We elaborate background about the structure of the FEWS NET data, as well as processing of these data to convert to a grid-cell format.

A1.1. Background

Fig. A1 summarizes temporal dimensions of the data based on FEWS NET's reporting cycles.

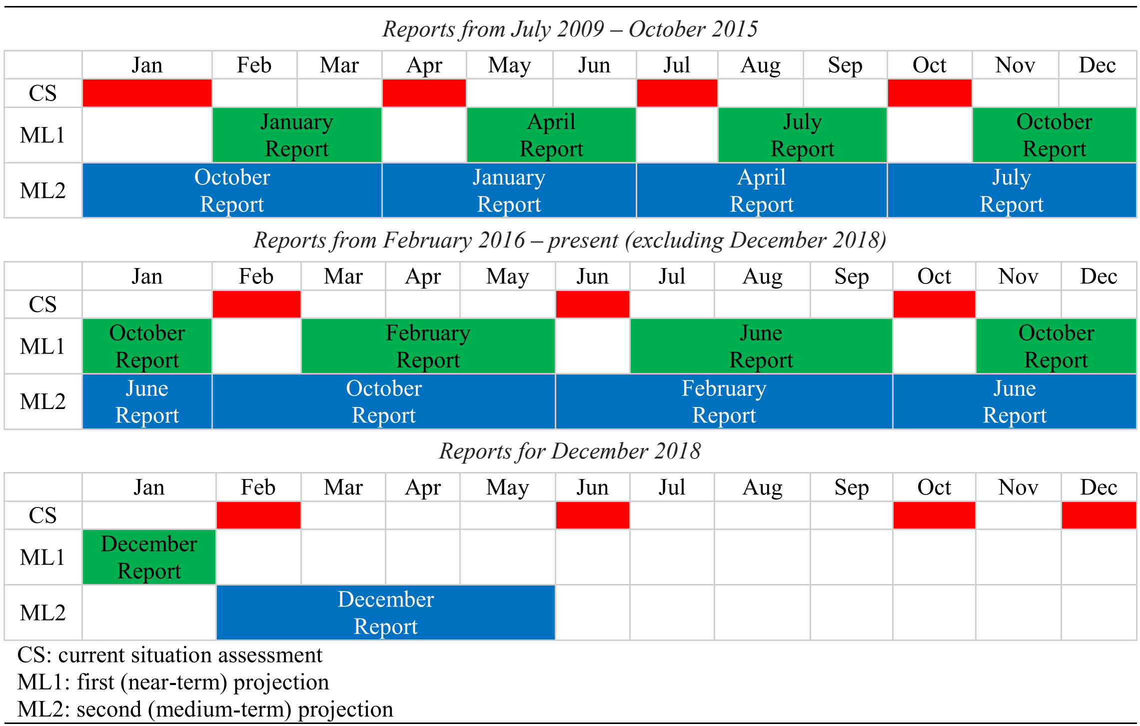

Fig. A1. Timing of FEWS NET assessments and projections. This graphic displays the standard structure of reporting cycles, which also defines the temporality and periodicity of the data. A feature is that the period covered by the medium-term projections (in blue) from a given report always overlaps with the month when the next report is released with updated current situation assessments (in red). The overlap provides a basis for calculating validation metrics. In contrast, the period covered by the near-term projections (in green) from a given report never overlaps with the month when the next report is released.

A1.2. Data processing

Current situation (CS) assessments and near-term (ML1) and medium-term (ML2) projections are each stored in separate shapefiles archived in FEWS NET's online Data Center. A shapefile includes a maximum of five (multi)polygons, differentiated by scores. For example, a single (multi) polygon denotes livelihood zones classified at level 1. (Multi)polygons can span subnational and country boundaries in addition to livelihood zones. Multipolygons – consisting of multiple non-contiguous polygons – are common.

With both polygons and multipolygons, the specific geographic units for which FEWS NET generated assessments and projections are not necessarily known. A given (multi)polygon may encompass one or more such assessments, and/or one or more projections. According to FEWS NET documentation, the most basic unit for generating assessments and projections is geographically disaggregated livelihood zones. Yet the (multi) polygons in FEWS NET's shapefiles of assessments and projections imperfectly map onto the (multi)polygons reflected in FEWS NET's separate livelihood zone shapefile (see http://fews.net/fews-data/335). Also, not all (multi)polygons remain static from report to report. Consequently, a given spatial unit from one report may lack a unit in the next report whose boundaries correspond exactly. The differences complicate comparisons across reports.

To address these considerations, we convert the shapefiles to a uniform, static grid of 0.5° × 0.5° cells whose dimensions correspond to around 55 km^2 at the Equator. The intent is to ensure that the spatial units for comparing ML2 projections to CS assessments are always consistent, since each grid-cell does not change boundaries over time. The grid-cells are also relatively similar in size, whereas the (multi)polygons and constituent livelihood zones within the FEWS NET shapefile data can vary substantially in size. The degree of uniformity among grid-cells enhances their comparability. For each report cycle, each grid-cell was assigned scores of the CS assessment and the ML2 projections, according to the raw shapefile data. A grid-cell entirely contained within the boundaries of a (multi)polygon in the raw data was assigned the scores for this corresponding (multi)polygon. A grid-cell that straddles the boundaries of more than one (multi)polygon in the raw data was assigned the scores for the (multi)polygon that constitutes the majority of the area of the grid-cell.

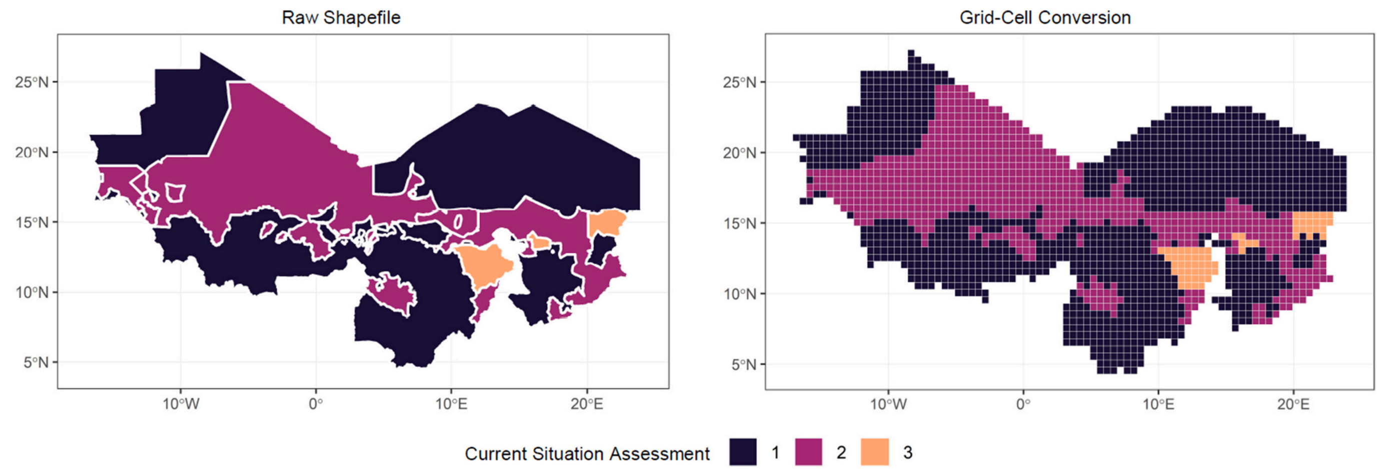

Fig. A2. Example of conversion of FEWS NET data to grid-cell format. These maps illustrate the conversion of current situation assessments from raw shapefiles in a format of irregular (multi)polygons to uniform grid-cells that are roughly comparable in size. The illustration uses data from FEWS NET's April 2014 Outlook Reports for West Africa.

Fig. A2 provides an illustrative example of the grid-cell conversion, using CS assessments included in the April 2014 reports for countries in West Africa. When we convert the irregular (multi)polygons shown in the left-hand panel to grid-cells as seen in the right-hand panel, a minor extent of the intricacy of the (multi)polygon boundaries is lost on the margins. Nonetheless, the granularity of the grid essentially captures the geographic variation in CS assessments that is evident in the raw shapefile data.

A2. Results

We present results of several distinct sets of supplementary analysis conducted with the following angles:

1. At a livelihood zone level.
2. At a grid-cell level weighting by population.
3. At a grid-cell level (unweighted) by region.
4. At a grid-cell level (unweighted) by county.
5. At a grid-cell level (unweighted) over time.

The first two sets of supplementary analysis are intended as robustness checks with respect to the main analysis reported in the article, which is conducted at a grid-cell level without weighting by population. Results of the remaining sets of analysis are mentioned in the article – and the graphical visualizations have been included here.

A2.1. Analysis by Livelihood Zone

We conduct analyses after overlaying FEWS NET's latest shapefile of livelihood zones on the data converted to a grid-cell format. Because multiple grid-cells can fall within each zone, we use modal ML2, CS, and humanitarian assistance flag values by zone for purposes of the analyses. The results are substantively similar to those obtained when conducting analysis at the grid-cell level.

Table A1
Validation metrics at the livelihood zone level. The sample includes 25 countries tracked by FEWS NET in Africa.

Full Sample, July 2009–June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	33.33%	-0.67	0.67	0.05%
4	30.62%	-0.83	0.83	1.37%
3	62.30%	-0.33	0.41	10.67%
2	73.23%	-0.01	0.27	26.80%
1	92.66%	0.08	0.08	61.11%
Overall	83.33%	0.00	0.18	100.00%
With Humanitarian Assistance Flag, April 2012 – June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	–	–	–	–
4	0.00%	-1.12	1.12	8.52%
3	36.79%	-0.64	0.64	26.57%
2	85.24%	-0.01	0.15	52.63%
1	73.47%	0.28	0.28	12.28%
Overall	63.66%	-0.24	0.38	100.00%

(*continued on next page*)

Table A1 (*continued*)

Full Sample, July 2009–June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
Without Humanitarian Assistance Flag, April 2012 – June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	–	–	–	–
4	33.33%	-0.86	0.86	0.64%
3	70.41%	-0.28	0.33	9.01%
2	75.76%	0.00	0.24	23.46%
1	92.84%	0.08	0.08	66.89%
Overall	86.43%	0.02	0.15	100.00%

A2.2. Population-Weighted Grid-Cell Level Analysis

Our main analysis at the grid-cell level treats all grid-cells with equal weight. The strengths of this approach are uniformity and simplicity in comparisons of units. A limitation is that population density is known to vary consequentially across areas of countries (and over time). Also, FEWS NET and stakeholders consider the caseloads of people affected by different levels of food security, not merely the geographic extent of area that is affected.

To capture this demographic dimension, we conduct a set of analyses in which grid-cells are weighted by population. For these purposes, we employ data from the Gridded Population of the World project (GPW v4.11; CIESIN, 2016). This resource supplies rasters of subnational population estimates around the world, which are derived using periodic census data for each country. Though the units in the GPW dataset have a high degree of spatial resolution (around 1 km^2), the estimates of population are actually only as granular as the administrative divisions reflected in each respective census. All spatial units within the same administrative division for a given release of the GPW dataset are assigned the same population estimate. Since 1995, an update of the GPW dataset has been released every five years, roughly corresponding to census waves around the world. GPW does not compile, impute, or interpolate data for intervening years between releases of updated data.

We match the GPW rasters to our $0.5^\circ \times 0.5^\circ$ grid and extract the GPW population estimate associated with each grid-cell. Our grid is less granular than the one used for the GPW rasters and therefore typically more congruent to the administrative divisions for which the GPW population estimates are derived. We employ the GPW datasets for 2005, 2010, 2015 and 2020, filling in intermediate years with the most recent estimate available (e.g., the GPW estimates from 2010 are applied to 2010–2014). We then rely on the population estimates as weights in the calculations of the validation metrics, with more populated grid-cells receiving proportionately greater weight than less populated cells.

As seen in Table A2 and Figs. A3-A4, weighting by population has only a marginal impact on the results. In particular, the performance in the weighted analysis is slightly better for the full sample, as well as the cases not flagged with humanitarian assistance, but slightly worse for the cases flagged with humanitarian assistance, relative to the unweighted analysis.

Table A2
Validation metrics at the grid-cell level weighted by population. The sample includes 25 countries tracked by FEWS NET in Africa.

Full Sample, July 2009–June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	50.93%	-0.64	0.64	0.08%
4	30.21%	-0.76	0.76	1.93%
3	65.38%	-0.33	0.37	11.18%
2	70.88%	-0.06	0.29	27.24%
1	93.52%	0.07	0.07	59.58%
Overall	85.71%	-0.001	0.15	100.00%
With Humanitarian Assistance Flags, April 2012 – June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	–	–	–	–
4	0.05%	-1.09	1.09	9.03%
3	38.67%	-0.63	0.63	26.38%
2	81.16%	-0.04	0.19	52.87%
1	71.32%	0.30	0.30	11.68%
Overall	57.04%	-0.31	0.45	100.00%
Without Humanitarian Assistance Flags, April 2012 – June 2020				
Level of Food Security for FEWS NET Medium-Term Projection	Accuracy	Bias	Absolute deviation	Share of cases
5	–	–	–	–
4	60.52%	-0.42	0.42	1.03%
3	78.44%	-0.19	0.23	9.21%
2	72.80%	-0.04	0.27	23.12%
1	94.49%	0.06	0.06	66.64%
Overall	89.50%	0.02	0.11	100.00%

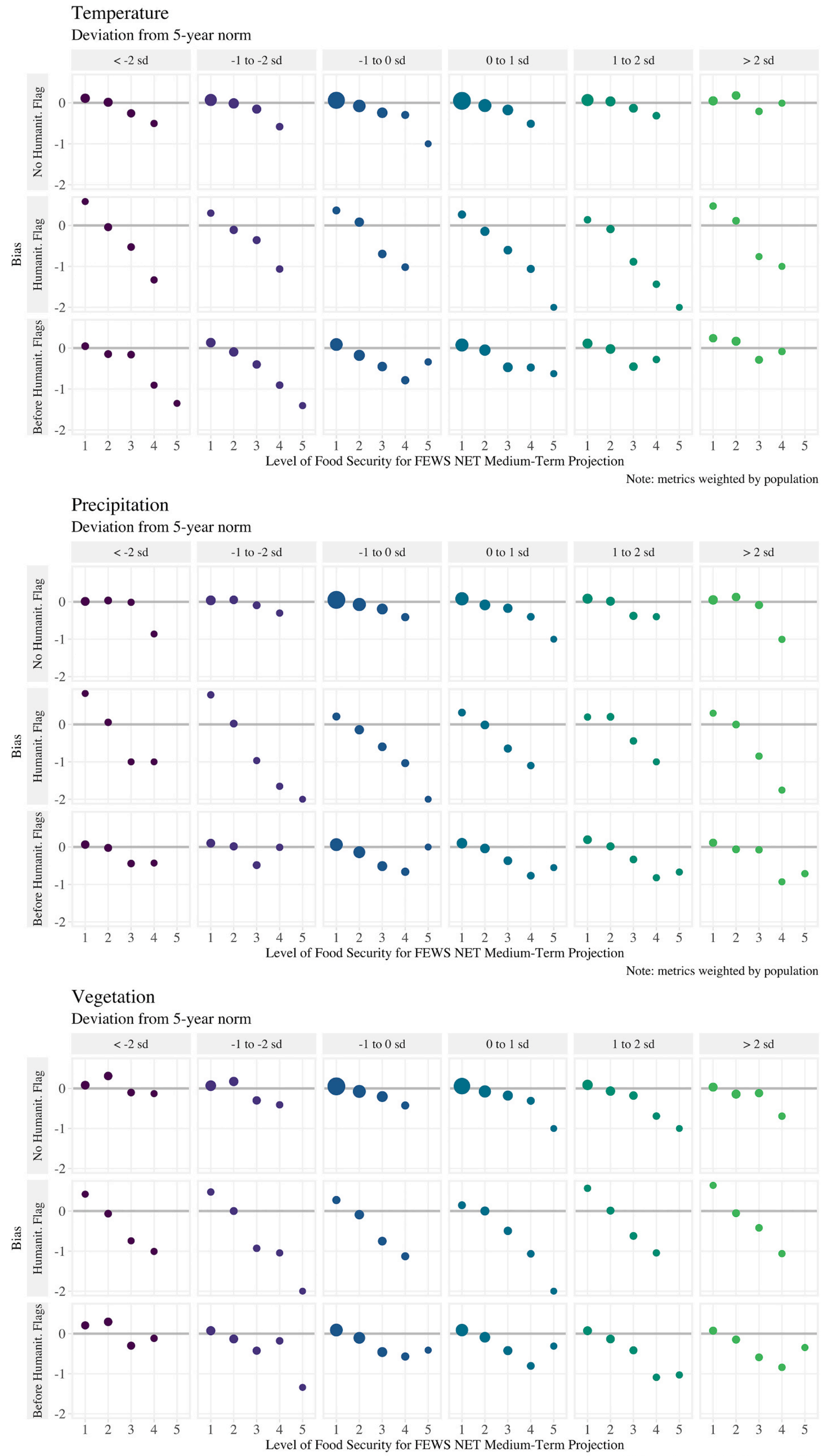

Fig. A3. Population-weighted bias in FEWS NET projections by variations in climatic conditions and humanitarian assistance. Standard deviations (sd) reflect the observed value of a climatic indicator for the month of the next FEWS NET report after a given projection, relative to the observed value of the indicator for the same month over the previous five years. The size of bubbles is proportionate to the share of grid-cell cases. All results span 25 countries tracked by FEWS NET in Africa. The top and middle sets of graphs in each panel cover April 2012–December 2018, whereas the bottom set of graphs in each panel covers July 2009–January 2012.

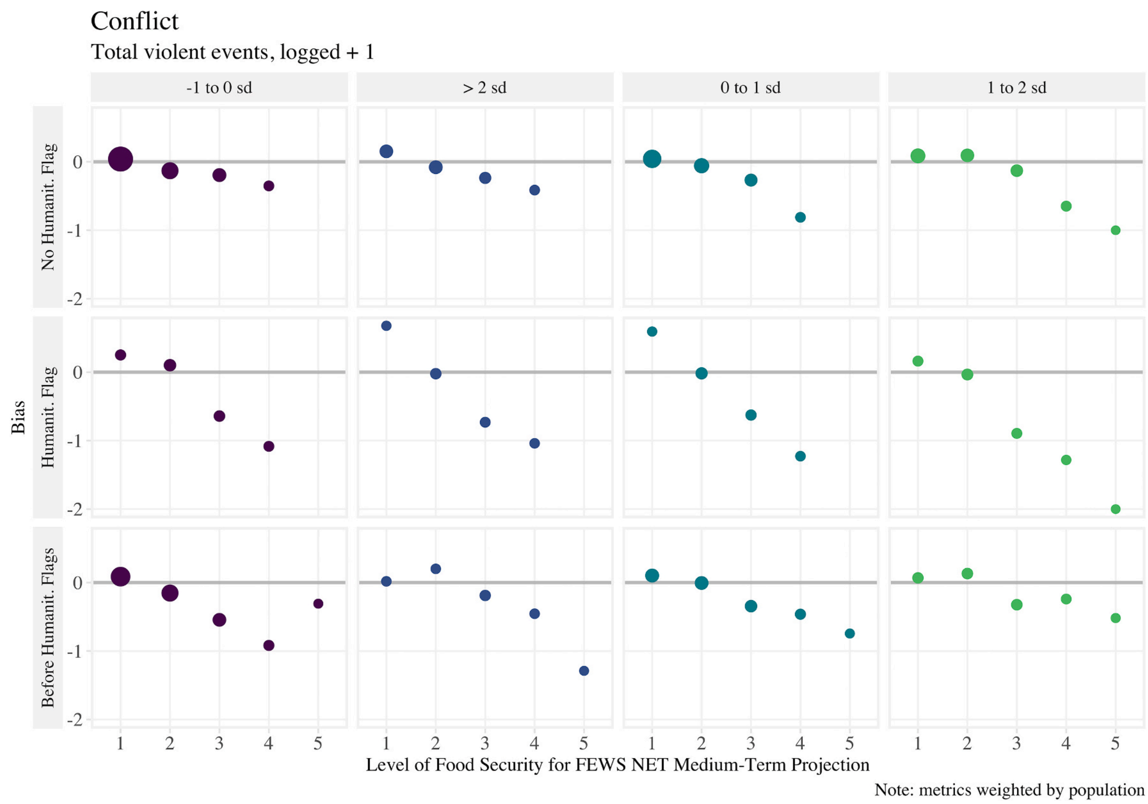

Fig. A4. Population-weighted mean bias by variation in violent conflict and humanitarian assistance. Standard deviations (sd) reflect the observed value of violent conflict events for the month of the next FEWS NET report after a given projection, relative to the observed values for the same month over the previous five years. The size of bubbles is proportionate to the share of grid-cell cases. All results span 25 countries tracked by FEWS NET in Africa. The top and middle sets of graphs cover April 2012–December 2018, whereas the bottom set of graphs covers July 2009–January 2012.

A2.3. Grid-Cell Level Analysis by Region

Fig. A5 summarizes the validation results aggregated across the countries within each of three regions of Africa. West Africa exhibits the highest accuracy and lowest bias and absolute deviation, while accuracy is lowest and bias and absolute deviation are highest for East Africa.

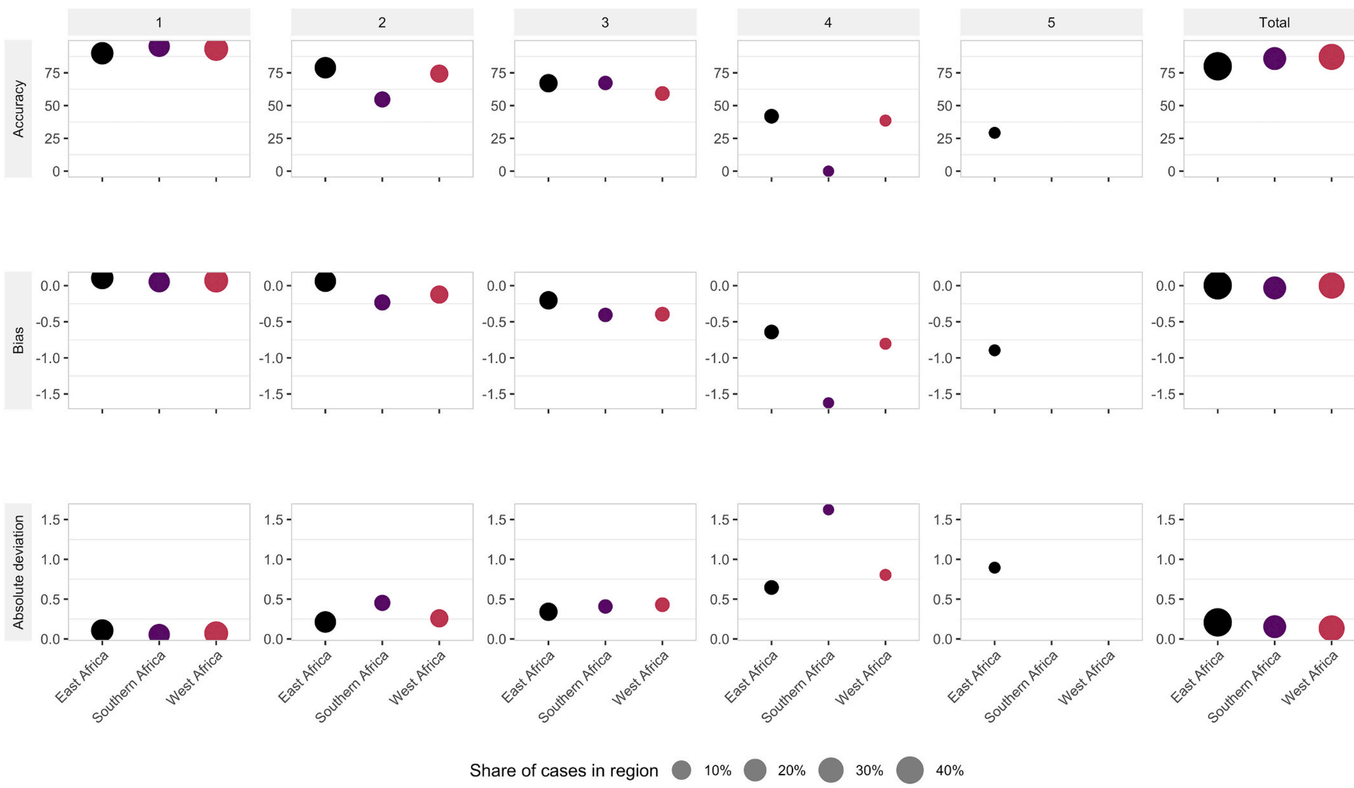

Fig. A5. Validation metrics by region. Each result reflects the mean for the validation metric across the entire sample of grid-cells, unweighted by population, within the countries in the region as defined by FEWS NET. The size of bubbles is proportionate to the share of grid-cell cases for the region. The sample includes 25 countries tracked by FEWS NET in Africa.

Further details underlying the accuracy of projections by region are presented in Table A3.

Table A3
Contingency table by region. This cross-tabulation presents the distribution of current situation assessment outcomes in the next FEWS NET report (in rows) relative to the corresponding medium-term projections in the previous FEWS NET report (in columns). Results are aggregated across grid-cells, unweighted by population, within the countries in each region as defined by FEWS NET. Each cell within the table displays the share of cases (top value) and number of cases (bottom value in brackets), by the level of the projection. Total cases by the level of projection are also listed at the bottom of each column. A case corresponds to a grid-cell unit for a given FEWS NET cycle. Darker shades of blue indicate higher shares of cases. The sample covers 25 countries tracked by FEWS NET in Africa from July 2009–June 2020.

East Africa

Level of Next FEWS NET Current Situation Assessment	1	2	3	4	5
5	0.00% [N=0]	0.00% [N=0]	0.17% [N=45]	0.20% [N=12]	29.21% [N=78]
4	0.00% [N=0]	0.15% [N=80]	6.58% [N=1694]	41.90% [N=2492]	52.06% [N=139]
3	0.50% [N=321]	13.45% [N=7177]	67.08% [N=17276]	51.86% [N=3084]	18.73% [N=50]
2	9.60% [N=6103]	78.90% [N=42107]	25.12% [N=6470]	5.58% [N=332]	0.00% [N=0]
1	89.90% [N=57179]	7.50% [N=4005]	1.05% [N=270]	0.45% [N=27]	0.00% [N=0]
	[N=63603]	[N=53369]	[N=25755]	[N=5947]	[N=267]

Level of FEWS NET Medium-Term Projection

Southern Africa

Level of Next FEWS NET Current Situation Assessment	1	2	3	4	5
5	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	[N=0]
4	0.00% [N=0]	0.57% [N=70]	0.16% [N=8]	0.00% [N=0]	[N=0]
3	0.75% [N=388]	11.04% [N=1351]	67.25% [N=3456]	37.66% [N=29]	[N=0]
2	4.05% [N=2096]	54.70% [N=6692]	24.58% [N=1263]	62.34% [N=48]	[N=0]
1	95.20% [N=49300]	34.26% [N=4191]	8.02% [N=412]	0.00% [N=0]	[N=0]
	[N=51784]	[N=12234]	[N=5139]	[N=77]	[N=0]

Level of FEWS NET Medium-Term Projection

West Africa

Level of Next FEWS NET Current Situation Assessment	1	2	3	4	5
5	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	0.00% [N=0]	[N=0]
4	0.00% [N=0]	0.28% [N=68]	2.10% [N=103]	38.59% [N=120]	[N=0]
3	0.26% [N=206]	6.24% [N=1500]	71.30% [N=3490]	42.44% [N=132]	[N=0]
2	6.55% [N=5288]	74.35% [N=17877]	44.37% [N=2172]	18.97% [N=59]	[N=0]
1	93.19% [N=75188]	19.13% [N=4599]	2.66% [N=130]	0.00% [N=0]	[N=0]
	[N=80682]	[N=24044]	[N=4895]	[N=311]	[N=0]

Level of FEWS NET Medium-Term Projection

A2.4. Grid-Cell Level Analysis by Country

Fig. A6 summarizes the validation results aggregated by country, grouped within each of three regions of Africa. The results for certain countries stand out as distinctive in these regions. Among the countries of East Africa, Kenya and Somalia exhibited the least accurate projections at level 1, implying that with some regularity FEWS NET assesses these countries as experiencing more serious outcomes despite expectations that food insecurity will be low. Across all countries in the region, the accuracy of projections at level 4 ranges between 25 and 50 percent. Among the countries of

Southern Africa, the Democratic Republic of Congo is the only country that tends toward under-projection of food insecurity at level 2. Across the region, accuracy of projections at level 4 likewise ranges between 25 and 50 percent. The select countries in West Africa with projections at level 4 present a contrast: accuracy was about 60 percent in Nigeria, whereas such projections in Mali and Niger were never accurate.

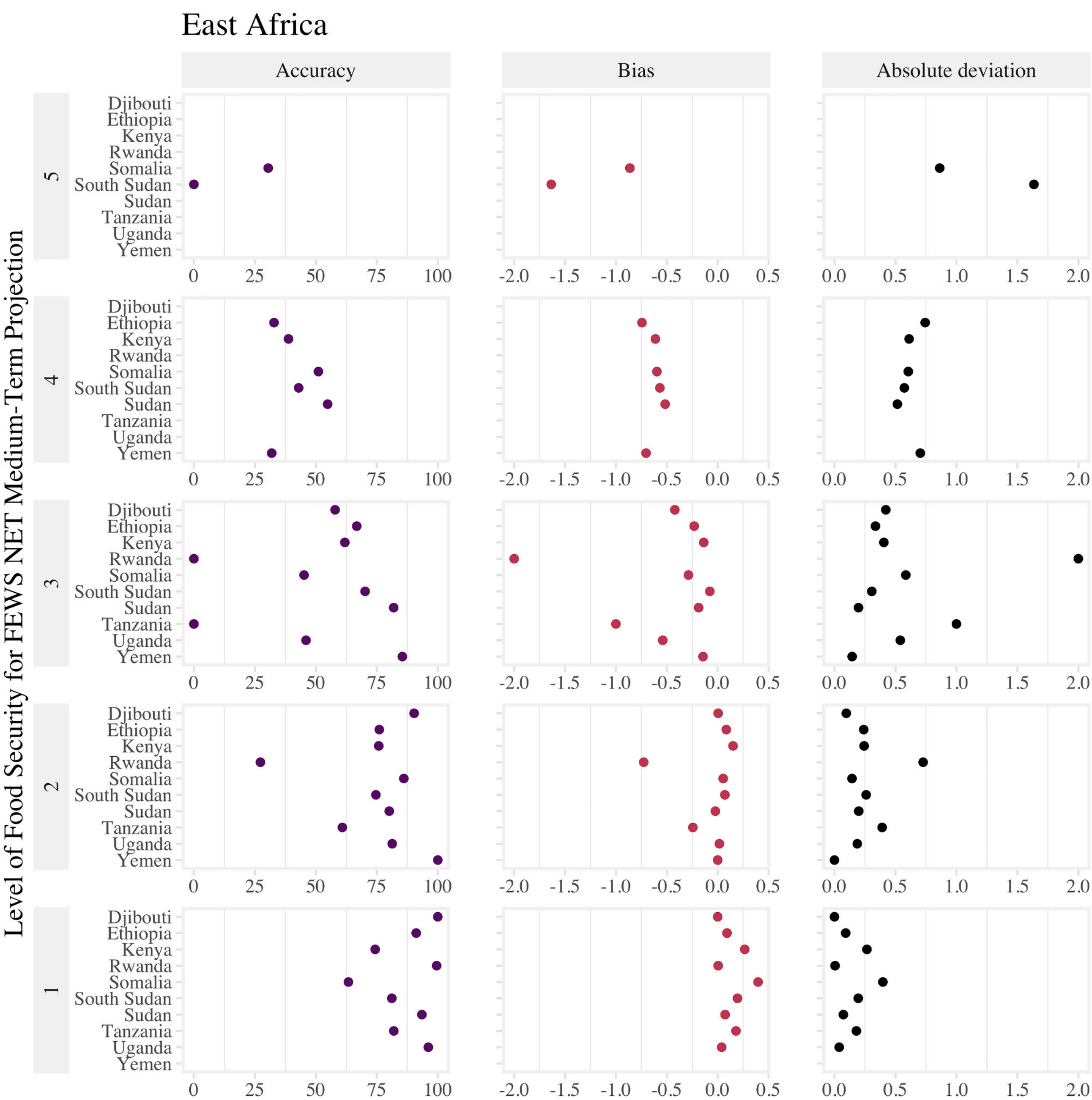

Fig. A6. Validation metrics by country. Each result reflects the mean for the validation metric across the entire sample of grid-cells, unweighted by population, within each country. Countries are grouped by regions as defined by FEWS NET. The sample covers from July 2009–June 2020. The levels are IPC compatible from April 2011 onwards.

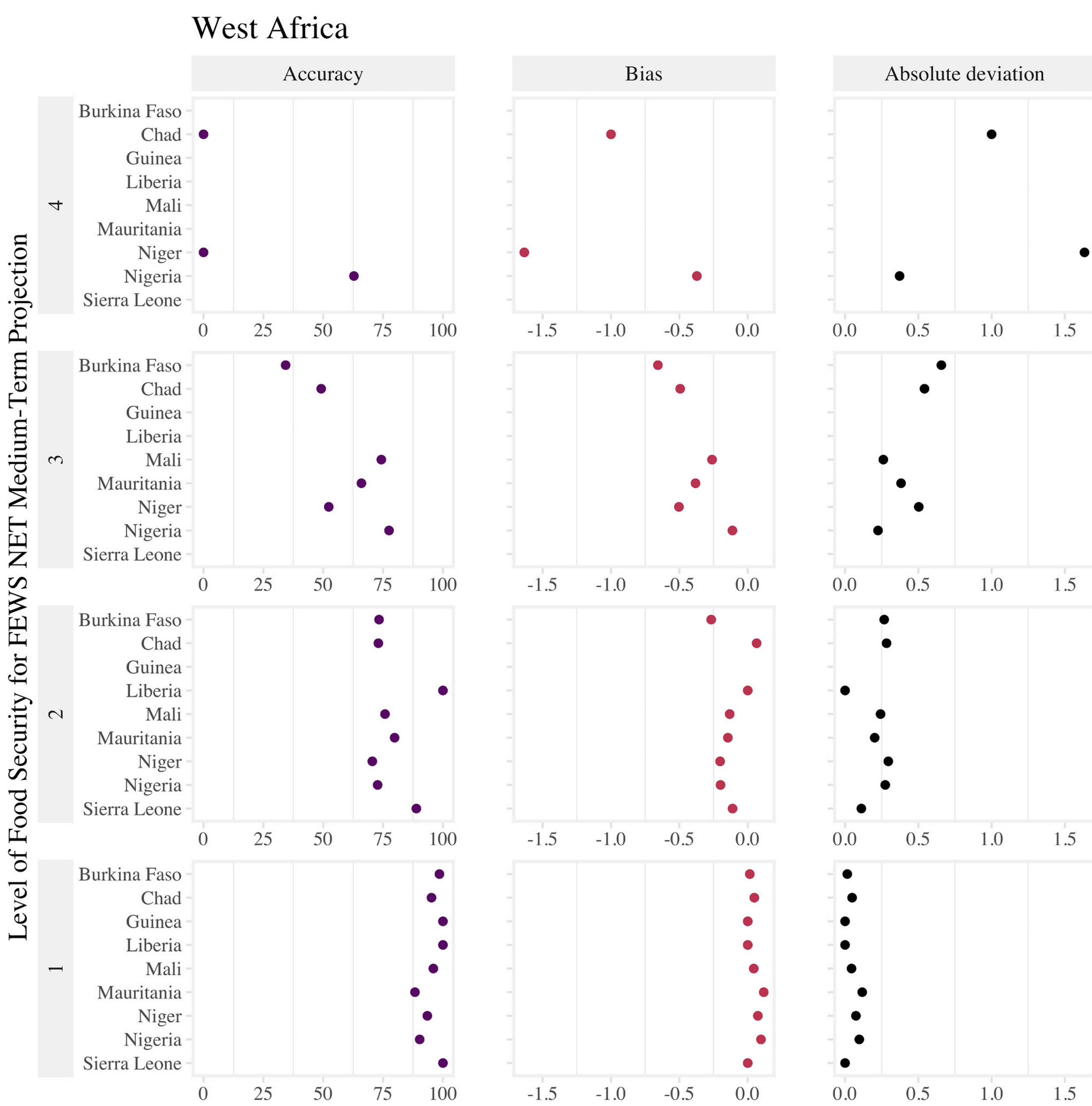

Fig. A6. (*continued*).

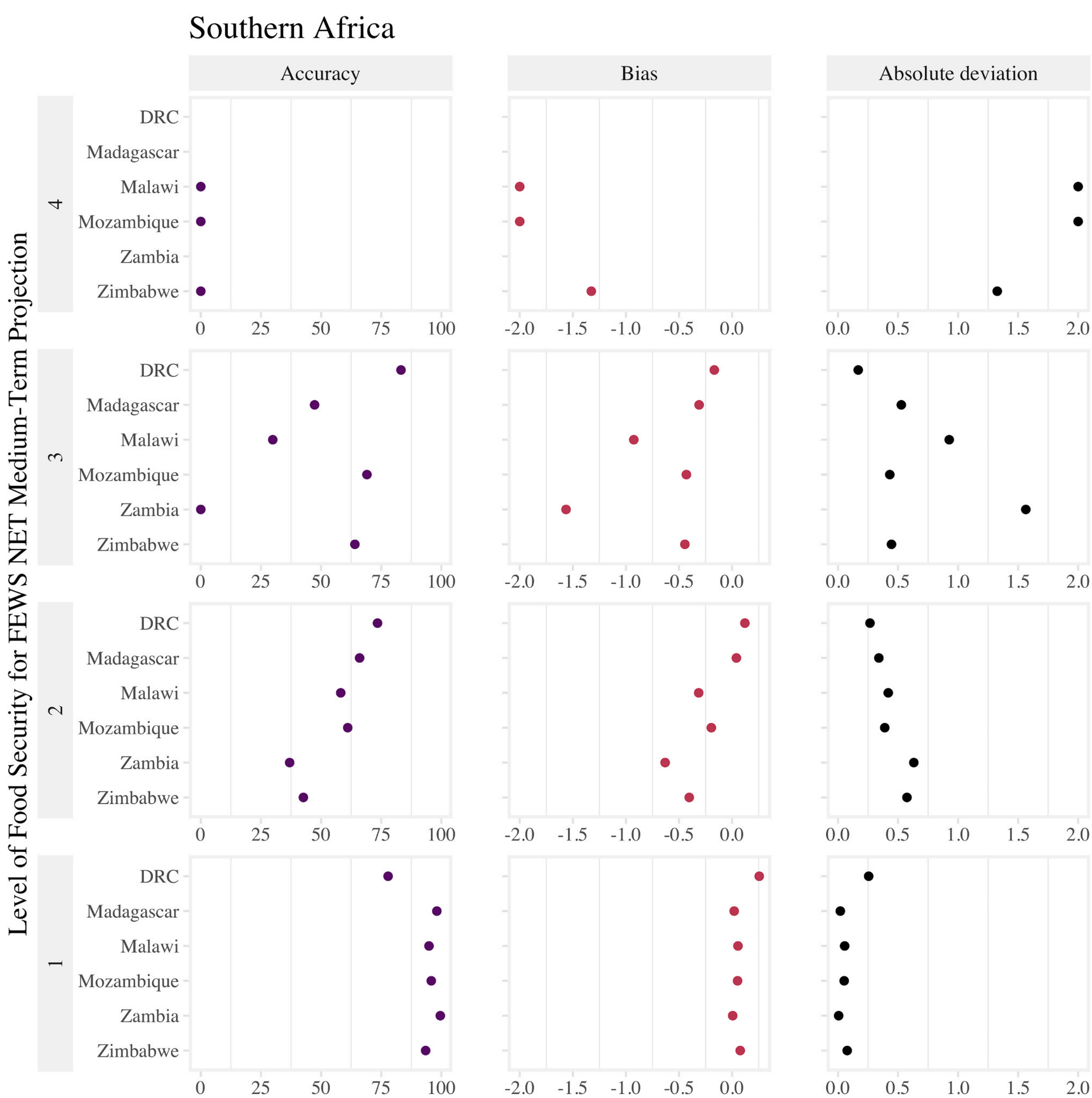

Fig. A6. (*continued*).

A2.5. Grid-Cell Level Analysis over Time

Fig. A7 presents time trends in the validation results for each level of projections. On balance, the results reveal declining accuracy, greater negative bias, and increasing absolute deviations, though the trends have not been monotonic. These trends are most pronounced for projections at level 4. Projections at level 3 exhibit the inverse pattern of improving accuracy and attenuating bias and absolute deviations.

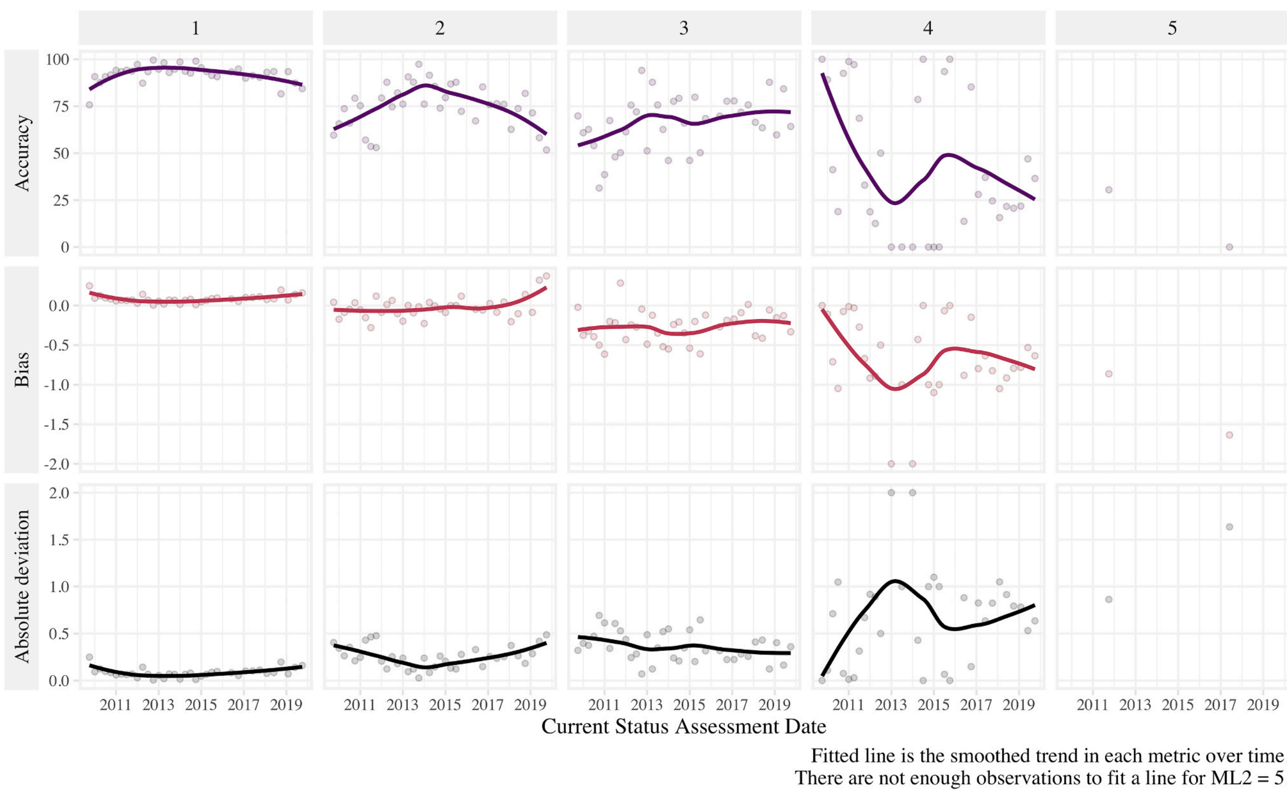

Fig. A7. Validation metrics over time. Results for the metrics are aggregated across the entire sample of grid-cells, unweighted by population, covering 25 countries tracked by FEWS NET in Africa from July 2009–June 2020.The fitted lines are smoothed trends in the metrics. Insufficient cases are available to fit lines for projections at level 5. Levels are IPC compatible from April 2011 onwards.

Fig. A8 displays an initial snapshot image from a time-lapse animation that maps the evolution of bias in FEWS NET medium-term projections by grid-cell.The animation links together the series of snapshots-one per reporting cycle-reflecting the entire analysis that covers 25 countries in Africa tracked by FEWS NET from July 2009 through June 2020. The user can pause, advance, rewind, and restart the animation. This visualization of the results, with the interactivity features, provides a tool for exploring and learning about geographic patterns and trends in the data on a comprehensive comparative basis, as well as for examining specific countries and time points in context.

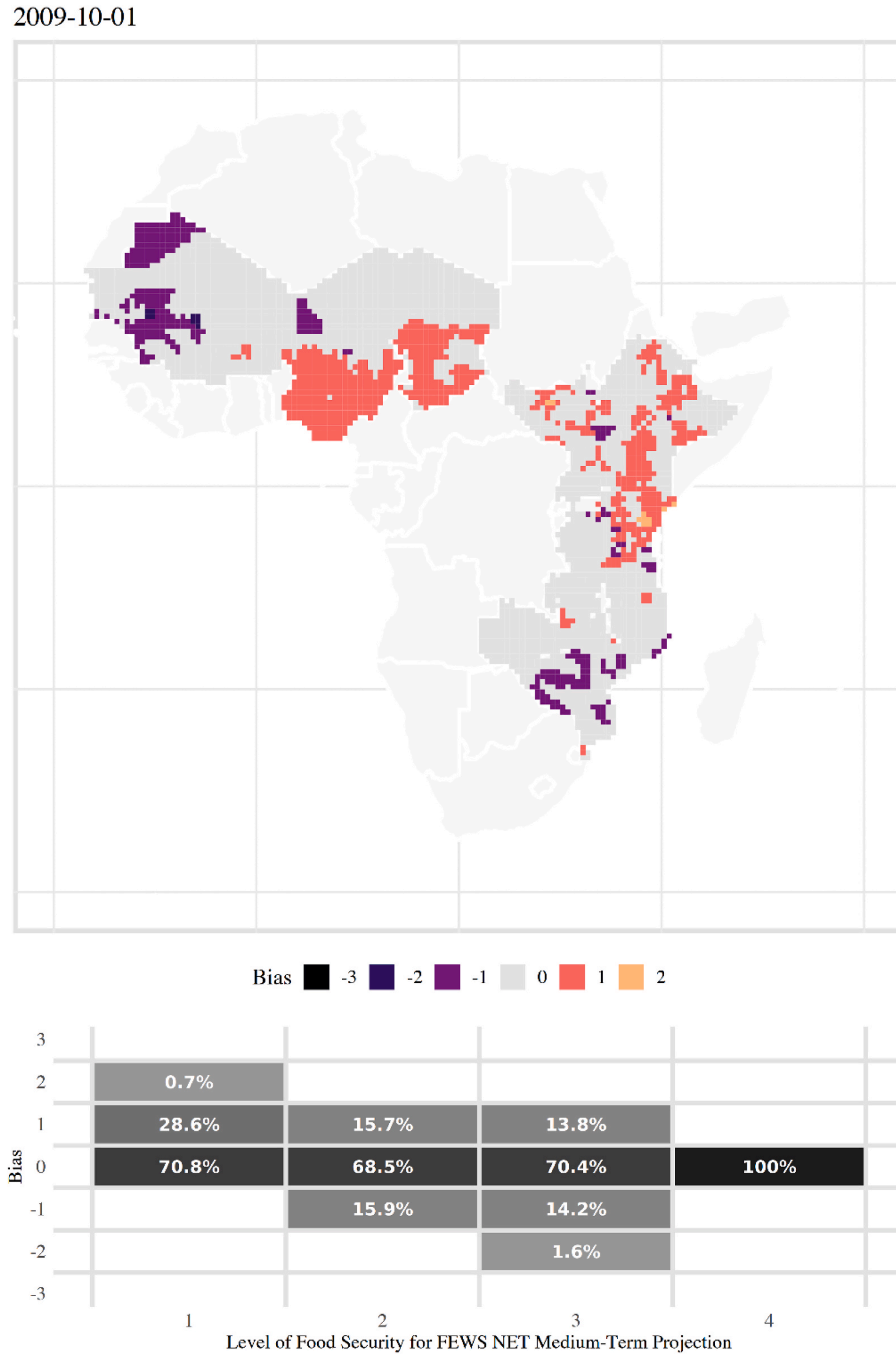

Fig. A8. Time-lapse animation of bias in FEWS NET medium-term projections of food security, by grid-cell.To access the animation, click on the URL https://dataverse.harvard.edu/api/access/datafile/4417531, which will launch a separate file archived on the Harvard Dataverse (Billing and Backer 2021). Bias is defined as the difference between the current situation assessment in a given report and the corresponding projection in the previous report. The values in the table are shares of grid-cells across the entire sample, by level of projection, that fall within each gradation of bias. The frames of the animation reflect the reporting cycles of FEWS NET and indicate the date of the report from which the current situation assessments were drawn.

References

Billing, Trey, 2021. Visualizing Trends in. Food Security across Africa, 2009–2020. Data and Animations at a Grid-Cell Level. https://doi.org/10.7910/DVN/LMSYG5. Harvard Dataverse, V3.

Brown, Molly E., 2008. Famine Early Warning Systems and Remote Sensing Data. Springer, New York, NY.

Brown, Molly E., Brickley, Elizabeth B., 2012. Evaluating the use of remote sensing data in the US agency for international development famine early warning systems Network. J. Appl. Remote Sens. 6 (1), 063511.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. Gridded population of the world, version 4 (GPWv4): population count. Palisades, NY: NASA socioeconomic data and applications center (SEDAC).

Choularton, Richard J., Krishnamurthy, P. Krishna, 2019. How accurate is food security early warning? Evaluation of FEWS NET accuracy in Ethiopia. Food Secur. 11 (2), 333–344.

Karsten, Donnay, et al., 2019. Integrating Conflict Event Data. J. Conflict Resolut. 63 (5), 1337–1364.

FAO, IFAD, UNICEF, WFP, WHO, 2019. The State of Food Security and Nutrition in the World 2019. Safeguarding against Economic Slowdowns and Downturns. FAO, Rome.

FEWS NET, 2018. Scenario Development for Food Security Early Warning, Guidance Document Number 1. https://fews.net/sites/default/files/documents/reports/Guidance_Document_Scenario_Development_2018.pdf. (Accessed 1 September 2020).

FEWS NET, 2018. Building Rainfall Assumptions for Scenario Development, Guidance Document Number 2. https://fews.net/sites/default/files/documents/reports/Guidance_Document_Rainfall_2018.pdf. (Accessed 1 September 2020).

Funk, Chris, et al., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. data 2 (1), 1–21.

Funk, Chris, et al., 2019. A high-resolution 1983–2016 T max climate data record based on infrared temperatures and stations by the climate hazard center. J. Clim. 32 (17), 5639–5658.

Krishnamurthy, P. Krishna, Choularton, Richard J., Kareiva, Peter, 2020. Dealing with uncertainty in famine predictions: how complex events affect food security early warning skill in the Greater Horn of Africa. Glob. Food Secur. 26, 100374.

Maxwell, Daniel, Hailey, Peter, 2020. The Politics of Information and Analysis. Humanitarian Information Systems and Analysis in Famines and Extreme Emergencies. Feinstein International Center, and Nairobi: Centre for Humanitarian Change, Medford MA.

National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2019. The Global Terrorism Database™. University of Maryland.

Raleigh, Clionadh, et al., 2010. Introducing ACLED: an armed conflict location and event dataset: special data feature. J. Peace Res. 47 (5), 651–660.

Ross, K.W., et al., 2009. Review of FEWS NET biophysical monitoring requirements. Environ. Res. Lett. 4 (2), 024009.

Sundberg, Ralph, Melander, Erik, 2013. Introducing the UCDP georeferenced event dataset. J. Peace Res. 50 (4), 523–532.

Verdin, James, et al., 2005. Climate science and famine early warning. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1463), 2155–2168.

Vermote, Eric; NOAA CDR Program, 2019. NOAA Climate Data Record (CDR) of AVHRR Normalized Difference Vegetation Index (NDVI), Version 5. NOAA National Centers for Environmental Information.