



OPEN On the forecastability of food insecurity

Pietro Foini¹, Michele Tizzoni^{1,2}, Giulia Martini³, Daniela Paolotti¹ & Elisa Omodei^{3,4}✉

Food insecurity, defined as the lack of physical or economic access to safe, nutritious and sufficient food, remains one of the main challenges included in the 2030 Agenda for Sustainable Development. Near real-time data on the food insecurity situation collected by international organizations such as the World Food Programme can be crucial to monitor and forecast time trends of insufficient food consumption levels in countries at risk. Here, using food consumption observations in combination with secondary data on conflict, extreme weather events and economic shocks, we build a forecasting model based on gradient boosted regression trees to create predictions on the evolution of insufficient food consumption trends up to 30 days in to the future in 6 countries (Burkina Faso, Cameroon, Mali, Nigeria, Syria and Yemen). Results show that the number of available historical observations is a key element for the forecasting model performance. Among the 6 countries studied in this work, for those with the longest food insecurity time series, that is Syria and Yemen, the proposed forecasting model allows to forecast the prevalence of people with insufficient food consumption up to 30 days into the future with higher accuracy than a naive approach based on the last measured prevalence only. The framework developed in this work could provide decision makers with a tool to assess how the food insecurity situation will evolve in the near future in countries at risk. Results clearly point to the added value of continuous near real-time data collection at sub-national level.

The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, calls for urgent action to “end hunger, achieve food security and improved nutrition and promote sustainable agriculture”¹. However, in 2019, 650 million people were still undernourished², with 135 million in 55 countries and territories reported to be acutely food insecure³. These numbers have significantly increased as a consequence of the COVID-19 pandemic, with at least 280 million people reported to be acutely food insecure in 2020⁴.

Food insecurity is a complex phenomenon, resulting from the interplay of many environmental, socio-demographic, and political events⁵. It is usually characterized through its four key pillars: availability, access, utilization, and stability; however, agency and sustainability have also been recently proposed as fundamental dimensions of food insecurity⁶.

Climate variability and extremes weather events are considered among the main drivers, mostly because of their effects on crop production^{7,8}. In recent years, studies have additionally shown that, beyond agricultural yields, extreme temperatures and precipitation conditions also directly negatively affect child nutrition and food security^{9–12}.

Conflicts are also deeply intertwined with food insecurity, of which they can be the cause or the consequence¹³. Several studies have shown that conflict impacts agricultural production, nutritional status, coping and consumption^{14,15}.

Food insecurity is also a global public health challenge. Household food insecurity is the leading risk factor of malnutrition, claiming approximately 300, 000 deaths each year. Whether directly or indirectly, due to inadequate food consumption and poor diet quality, it is also accountable for over half of all deaths among children in Sub-Saharan Africa. Among the various determinants of food insecurity in relation to child malnutrition, the main ones are socio-demographic characteristics as well as food prices¹⁶. The impact of food insecurity on health cannot be overstated^{17–19}. The current COVID-19 crisis has exacerbated the situation and undermined some of the critical elements influencing food insecurity, e.g. international agricultural supply chains and goods prices impacted by trade restrictions²⁰. In low and middle income countries, social distancing, workplace closures, and restrictions on mobility and trade had cascading effects on economic activity, food prices, and employment²¹.

From a policy perspective, food insecurity will remain a worldwide concern for the next 50 years and beyond. As a 20-year old article in Science stated: “Recently, crop yield has fallen in many areas because of declining

¹ISI Foundation, Via Chisola 5, 10126 Turin, Italy. ²Department of Sociology and Social Research, University of Trento, Via Verdi, 26, 38122 Trento, Italy. ³World Food Programme, Research, Assessment and Monitoring Division (RAM), Via Cesare Giulio Viola 68, 00148 Rome, Italy. ⁴Department of Network and Data Science, Central European University, Quellenstraße 51, 1100 Vienna, Austria. ✉email: omodeie@ceu.edu

investments in research and infrastructure, as well as increasing water scarcity. Climate change and HIV/AIDS are also crucial factors affecting food security in many regions. Although agroecological approaches offer some promise for improving yields, food security in developing countries could be substantially improved by increased investment and policy reforms²². In the past twenty years, the situation has however significantly worsened and agroecological technologies are not keeping up with the accelerating effects of climate change that, ever increasingly, are affecting countries and populations on much shorter scales (weeks and months rather than years) than in the past.

Therefore, an essential step towards achieving hunger reduction is to have access to frequent, up-to-date information on the status of food insecurity in countries facing humanitarian crises, and to estimates of where and when the situation is likely to improve or deteriorate, in order to allow for informed and timely decision-making on resource allocation and on relevant policies and programmes. For this reason, food security assessments are performed on a regular basis. This is done through face-to-face surveys as well as computer-assisted telephone interviews (CATI), which have become very popular in the last few years, and are sometimes complemented by other technologies like interactive voice response and web surveys. The World Food Programme (WFP) is currently monitoring the food security situation in near-real time in a number of countries at the sub-national level, collecting on a daily basis, through remote phone surveys, information on levels of food consumption and food-based coping, as well as other relevant indicators^{23,24}. This unprecedented availability of daily sub-national level data paves the way for new possibilities since it not only allows for a continuous up-to-date picture of the current situation, but could also be used to build predictive models to forecast how the situation will evolve in the future. In this study, we explore the forecastability of insufficient food consumption levels, and show, specifically for Syria and Yemen, that satisfactory predictions up to 30 days into the future can be obtained when enough daily sub-national level historical data is available. Having access to regular real-time estimation of how the situation is likely to evolve in the near future would allow WFP for more informed discussions on need-based humanitarian assistance allocation decisions.

Forecasting modeling has been the subject of extensive investigation during the last decade in different fields, from financial markets^{25,26} to infectious disease epidemiology^{27–30}. However, it is still a relatively new area of research in the context of food security. The Food and Agriculture Organization of the United Nations (FAO) developed a methodology to produce annual country-level estimates of the prevalence of undernourishment, and to project these estimates up to 10 years into the future^{31,32}. Okori and collaborators first proposed to use machine learning models to predict whether a household is in famine or not from household socioeconomic and agricultural production characteristics^{33,34}. The effort of predicting levels of insufficient food consumption has been tackled in the context of Malawi, in a study where the authors built a model trained on 2011 data to estimate the situation in 2013³⁵, and more recently in a work proposing a model to nowcast sub-national levels of insufficient food consumption on a global scale³⁶. Both studies propose methods to predict the current situation when primary data is not available, but they do not address the challenge of making projections for the future. The World Bank recently proposed a machine learning approach to forecast transitions into critical states of food insecurity³⁷ and a stochastic model to forecast famine risk³⁸. Concurrently, similar work was carried out by other researchers in the context of Ethiopia³⁹. These studies focus on forecasting month-to-month transitions to different phases of food insecurity, based on the Integrated Food Security Phase Classification (IPC) framework⁴⁰.

In this study, we tackle a different problem: forecasting the daily evolution of the prevalence of people with insufficient food consumption at the sub-national level. This metric, characterizing a given area at a given time, is defined as the prevalence of households, in the specified area and time period, that are identified to have poor or borderline food consumption. Such prevalence is computed from a representative number of household surveys enquiring about one of the core household food insecurity indicators, namely the Food Consumption Score (FCS), which captures households' dietary diversity and nutrient intake⁴¹. The FCS is one of the many indicators developed to monitor food insecurity, each capturing different dimensions of the problem. Other examples include the Household Dietary Diversity Score (HDDS), also focusing on dietary intake, the Coping Strategies Index (CSI) and the Household Hunger Scale (HHS), both focusing on the consequences of constrained access to food⁴², and the more recent Food Insecurity Experience Scale (FIES), focusing on behaviors and experiences associated with difficulties in accessing food due to resource constraints⁴³. In this study, which constitutes only a first attempt at food insecurity daily forecasting, we choose to focus on the FCS, given its wide adoption by WFP and the consequent data availability.

Having access to reliable predictions of the evolution of insufficient food consumption levels over future weeks and months could allow governments and organizations to identify which areas should be monitored more closely and to eventually take timely decisions on resource allocation. Hence, the goal of this study is to develop a forecasting model able to predict, in countries with major food crises, the daily sub-national prevalence of people with insufficient food consumption up to 30 days into the future.

The main difference with IPC and the Famine Early Warning Systems Network (FEWS NET)⁴⁴ is that, while these make use of consensus-based expert opinion, our approach is algorithmic and data-driven, and can hence be applied in an automatic fashion. IPC's and FEWS NET's projections are essential for humanitarian action, however they require local expertise and considerable time to be developed. Our main goal is not to replace these efforts, but rather to complement them with an approach that, once improved as more data becomes available, could be used to provide rapidly available forecasts for several places at the time by automatically feeding near real-time data to the proposed algorithms.

Results

Time trends of insufficient food consumption. We study the possibility of forecasting one of the core dimensions of food insecurity by means of a unique data set of daily sub-national time series of the prevalence of people with insufficient food consumption, in six countries in West Africa and the Middle East: Burkina Faso, Cameroon, Mali, Nigeria, Syria and Yemen (see the Methods for a detailed definition of the indicator under investigation). These countries, although having varying socioeconomic and geopolitical characteristics, have all been identified as major food crises where acute food insecurity is driven by conflict, weather extremes and economic shocks³. Among all major food crises, these are the countries for which the largest volume of food insecurity data is available.

The length of these time series varies from a minimum of 865 days in Mali 1340 days in Yemen, over the years 2018–2022. Also, geographic coverage varies across countries. In the case of Burkina Faso, the prevalence of insufficient food consumption is available for all administrative units of the country, while only 3 states of Nigeria are included in our dataset (Adamawa, Borno, Yobe), given these are the most at risk areas closely monitored by WFP. Overall, our dataset covers 88% or more of the total population in all countries, with the only exception of Nigeria (see Supplementary Table S4).

Since the mode of questionnaire administration can have serious effects on data quality^{45,46}, during the last few years WFP conducted mode experiments in several countries, each demonstrating the feasibility of collecting food security indicators via CATI surveys^{47,48}. However, sampling and selection bias should be assessed and mitigated^{49–51}, hence post-stratification weights were applied by WFP, as detailed in the Methods section.

As shown in Fig. 1, in the six countries, time trends of insufficient food consumption display noisy and irregular patterns, underscoring the complex dynamics underlying food insecurity. During the study period, all countries experienced large fluctuations in the prevalence of insufficient food consumption, and such variations were not uniform between sub-national administrative units. In Cameroon, for instance, only a few regions were characterized by a relatively high proportion of food insecure people, generally above 50%, but also exhibiting large fluctuations, such as the rapid decline and subsequent increase observed in the North-West regions. On the other hand, in Syria, the sub-national trends were all similar in terms of relative changes in the affected population, with a general upward trend affecting almost every province beginning in July 2020. In the governorates of Yemen, for which the longest time series are available, the proportion of the population affected by food insecurity varied between 20% and 60% during the years 2018–2022, however, a common national time trend is less recognizable. It should be noted that some of these irregularities could also be partially due to the effects of sampling and selection bias, whose mitigation can rarely be achieved in full.

Permutation entropy and intrinsic predictability of food insecurity. We first quantify the intrinsic predictability of the time series shown in Fig. 1 by means of a permutation entropy analysis. Permutation entropy (PE) is a model-free measure of time series complexity^{52,53}, that is conceptually similar to the Shannon entropy but is based on the frequency distribution of motifs. PE has been extensively used to assess the predictability of time series in different domains including finance and economics^{54,55}, ecology⁵⁶ and infectious disease epidemiology³⁰. In short, to compute the PE of a time series we translate its real valued sequence (x_1, x_2, \dots, x_N) into a frequency distribution of symbols that represent patterns of relations $x_i < x_j$, $x_i = x_j$ or $x_i > x_j$ between nearest or distant neighbors, x_i and x_j . Such frequency distribution is then used to assess the predictability of the time series by computing the Shannon entropy associated with the distribution of permutation patterns in the symbols defined above. In the Methods section we provide a complete formal definition of the PE and its computation. It has been shown that PE can be considered as a measure of intrinsic predictability of a time series and its value is positively associated with forecasting error⁵⁶. Intuitively, PE quantifies the information that is transmitted from the past to the present state of a time series: a time series that periodically visits the same few symbols among the many possible will have a low entropy and its present state will be easily determined from the past. A random time series that uniformly samples the symbols with equal probability will have a high entropy and its future will not be predictable from past states.

In the case of food insecurity, we find that insufficient food consumption trends are not easily predictable based on their past history. As shown in Fig. 2, their predictability, measured as $\chi = 1 - H$, where H is the PE, never reaches values above 0.5 and it is often reduced to 0.1–0.2 within a 10-day horizon. These values are generally much lower than those observed in the case of infectious disease dynamics³⁰ and they are closer to measures of predictability of financial time series⁵⁷, which are characterized by a high short- and long-term volatility. Confidence intervals around mean predictability values are also narrow, highlighting a consistent lack of recurrent patterns in the insufficient food consumption time series across different time scales, which in turn highlights the presence of intrinsic entropy barriers to their predictability.

Forecasting food insecurity with secondary information. Following from the observation that insufficient food consumption trends are not highly predictable from their own history, we explore whether secondary information can be used to enhance our ability to predict their future evolution. To this end, we revert to information on the key drivers of food insecurity: conflict/physical insecurity, extreme weather events and economic shocks⁵⁸. We build a set of indicators covering these three domains and develop a forecasting model based on gradient boosted regression trees (XGBoost)⁵⁹ to make predictions on how the insufficient food consumption trend will evolve up to 30 days into the future. More specifically, in our model we consider as predictors of insufficient food consumption the following indicators (see Methods and the Supplementary Information file for a full description). First, we include daily time series of the prevalence of people using crisis or above crisis food-based coping, which is obtained from another core food insecurity indicator, the reduced Coping Strategy Index (rCSI), by measuring the share of households with $rCSI \geq 19$ ^{40,60}. Since political unrest

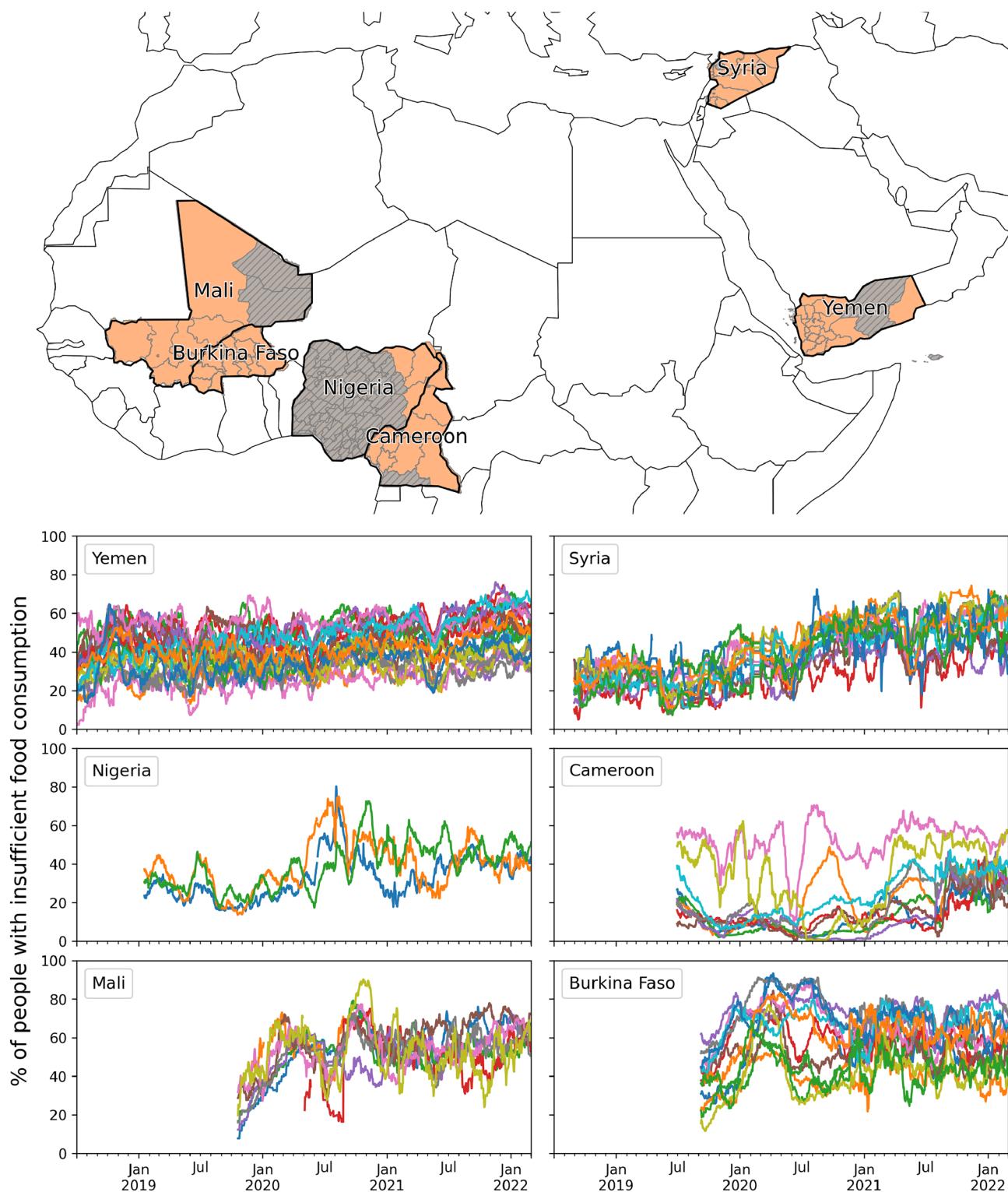


Figure 1. Time trends of insufficient food consumption. Each panel displays daily time series of the percentage of people with insufficient food consumption in the first-level administrative units of Burkina Faso, Cameroon, Mali, Nigeria, Syria and Yemen. The six countries are highlighted in the map, and the orange shade indicates the areas that are considered by our analysis. The map was created by the authors using GeoPandas v0.9.0⁷⁹ and the shapefiles provided by OCHA under the CC BY-IGO license on HDX (<https://data.humdata.org/>).

can affect food security, we include in our model daily time series of fatalities due to conflict or political violence as reported by the Armed Conflict Location and Event Data Project (ACLED)⁶¹. Economic shocks are included

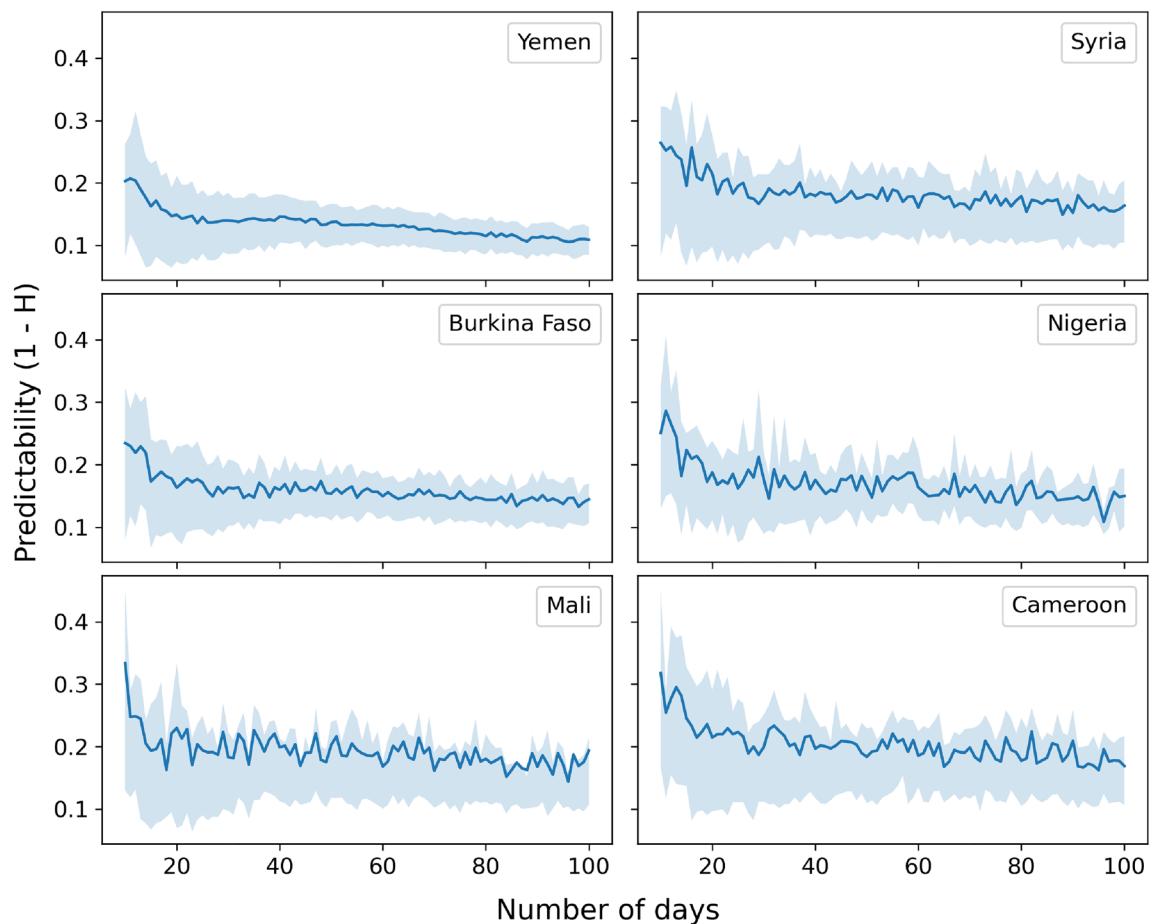


Figure 2. Food insecurity is characterized by low predictability. The average predictability $\chi = 1 - H$ for daily trends of insufficient food consumption, in the six analyzed countries, is shown as a function of time series length in days. We average H over temporal windows by selecting 1000 random points from each time series and calculating H for windows of length 10, 11, 12, ..., 100 days. The solid lines indicate the mean value and the shaded areas mark the interquartile range across all administrative units and starting locations in the time series.

into the model by considering monthly variations in the price of cereals and tubers in local currency. The model takes into account the effects of weather events and climate conditions by including time series of rainfall, of its anomaly with respect to long-term averages (over 1 and 3 months), and time series of the Normalized Difference Vegetation Index (NDVI), a standard satellite-based measure of vegetation coverage that is commonly used for drought assessment⁶², and of NDVI anomaly. Finally, since the food consumption behavior of most of the population in several African and Asian countries is affected by Ramadan, we include a time series that marks the days of the Ramadan period that fall within the time window used to measure people's food consumption.

Figure 3 and Supplementary Fig. S1 show the prediction results of the model for the case of Yemen and Syria, respectively, the countries for which the longest time series of insufficient food consumption prevalence are available. In Yemen (Syria), cross-validated predictions can explain between 99% (99%) and 72% (47%) of the variation in insufficient food consumption, with the former being the variation explained by the 1-day into the future forecast, and the latter for the 30-day into the future one (Fig. 3a and Supp. Fig. S1a). This is a significant increase of R^2 with respect to a naive prediction based on the last measured value only, which can only explain between 99% (99%) and 65% (31%) of the variation (Fig. 3c and Supp. Fig. S1c), and whose mean squared error (MSE) is larger and with a wider dispersion than the MSE of the proposed model (Fig. 3b, d and Supp. Fig. S1b, d). Specifically, panels c and d show that the proposed model clearly outperforms the naive approach for high values of the prediction horizon, consistently with the expectation that as time progresses the last measured value is not a good guess anymore. The scatterplots in Fig. 3e and Supp. Fig. S1e show the performance of the forecasting models as the predicted insufficient food consumption value against the actual value, for different prediction horizons. As expected, dots get further away from the identity diagonal as the prediction horizon increases up to 30 days, although the general behavior is consistent with a good predictive accuracy.

Over short forecasting horizons, typically less than 14 days, a naive approach proves to be a good enough predictor as we do not expect food consumption to suddenly change from one day to the next. However, as we try to forecast further into the future, we see that the forecasting model starts to outperform the naive approach, as shown in Fig. 3f (Supp. Fig. S1f) for the case of two Yemeni (Syrian) provinces over a 30 day horizon.

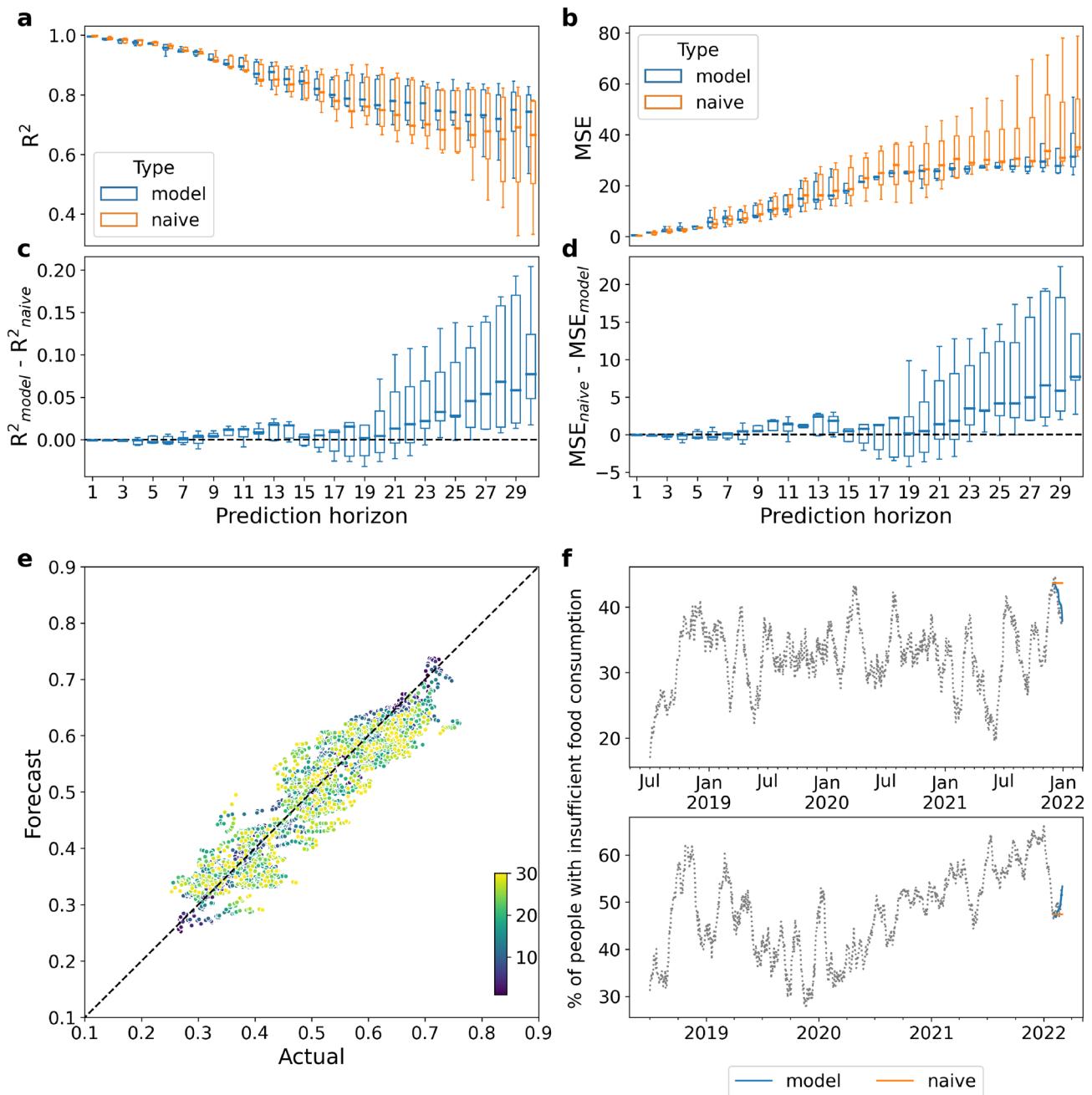


Figure 3. Forecasting the prevalence of people with insufficient food consumption in Yemen. The forecasting is performed over 5 different monthly splits of all governorates time series, from October 2021 to February 2022. **(a)** Box plots of the coefficient of determinations (R^2) across the 5 splits for both the proposed and the naive models (in blue and orange, respectively), for each forecasting horizon. **(b)** Box plots of the mean squared error (MSE) across the 5 splits for both the proposed and the naive models for each forecasting horizon. **(c)** Box plots of the difference between the R^2 of the proposed and of the naive model for each split. **(d)** Box plots of the difference between the MSE of the naive and of the proposed model for each split. **(e)** Predicted vs actual value for each data point in the 5 splits. Colors represent the corresponding forecasting horizon and vary from dark blue (1 day) to yellow (30 days). **(f)** Example of forecasting results for December 2021 in Amanat Al Asimah (top) and February 2022 in Abyan (bottom).

In the case of the remaining four countries, whose available training points are less than half than for Yemen, the model performance is worse than the naive approach across all prediction horizons, both in terms of R^2 and MSE (see Supplementary Figs. S2–S5).

Model performance as a function of data availability. Given the relatively poor predictive performance of our models in countries with short time series of insufficient food consumption, we systematically

examine how the performance varies as a function of the length of the time series available to train the model and of spatial coverage, indicating the number of sub-national areas. We find that, compared to the naive approach, the performance of our model dramatically increases with the number of available training points, which is given by the product of the two dimensions above: temporal length and spatial coverage (see Fig. 4). Moreover, with a given size of the training set, the proposed model tends to perform better than the naive approach as the forecasting horizon grows, demonstrating that, as expected, the model is better at predicting further into the future than just considering the last available measurement. However, this effect is evident only when a large training set is available (as in the case of Yemen, with more than 20,000 data points), and a small training set reduces the benefit of the model even over longer time horizons.

A limitation of this analysis relies on the fact that in our dataset different numbers of training points coincide with different countries, which does not allow to disentangle effects due to the local context from those due to temporal length and spatial coverage. We therefore performed further analyses in order to separately investigate the role played by the number of covered areas and by the time series length. First, we considered the same time series length for all countries by using as starting date for all of them the earliest date available for all countries. In this setting, the difference in the number of training points among the different countries is only due to the different number of areas covered. Results are reported in Supplementary Fig. S6a. Secondly, we considered instead the full time series but we fixed the number of considered areas to the minimum number of areas available for all countries (in this case we had to exclude Nigeria since only three areas are covered there, which would have been too few for the analysis). In this setting, the difference in the number of training points among the different countries is only due to different lengths of the available time series. Results are reported in Supplementary Fig. S6b. In both cases, we re-trained and tested the model using, each time, a different data subset, as described above. Results show that the performance of the model still increases with the dimension under study (the length of the time series and the number of covered areas), confirming our initial results.

Discussion

In this study we addressed the critical challenge of forecasting the daily evolution of a food security indicator, namely the prevalence of people with insufficient food consumption, as measured by WFP. The problem promptly proved difficult given that the analyzed time series exhibit noisy and irregular behavior. This is to be expected since food insecurity in Sub-Saharan Africa and the Middle East is a highly dynamic phenomenon, comprising a seasonal component related to agricultural production calendars and religious observances such as Ramadan (during which consumption patterns are completely altered), but also subject to swift changes when external shocks hit, such as the emergence of conflict, extreme weather events or economic shocks^{63–65}.

Therefore, forecasting based solely on information on the historical evolution of the target indicator over time would not be successful, as demonstrated through a permutation entropy analysis. Hence, we extended the proposed framework to include historical information on the key drivers of food insecurity and built models that comprise both endogenous (insufficient food consumption itself, as well as food-based coping information) and exogenous factors (conflict-related fatalities, rainfall and vegetation and their anomalies, staple food prices and Ramadan's occurrence). We showed that the proposed model makes it possible to forecast the prevalence

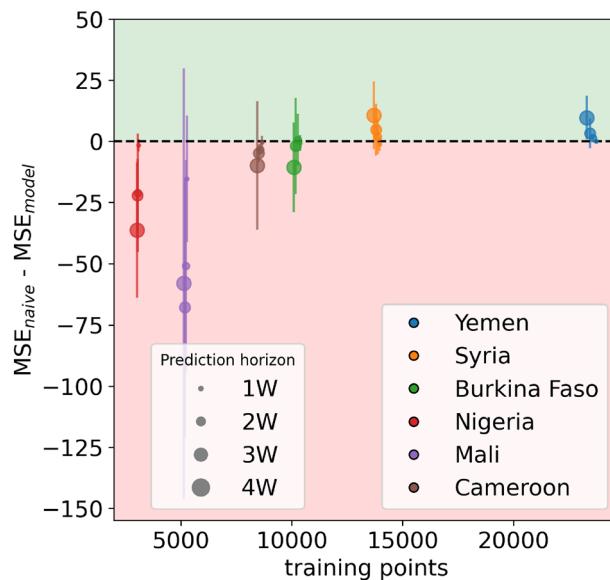


Figure 4. Model performance as a function of the number of available training points. For the six analyzed countries and four different forecasting horizons (1–4 weeks), the figure shows the averaged differences between the MSE of the naive approach and the MSE of the forecasting model across the different splits, as a function of the size of the training set. Error bars correspond to the relative standard deviation. The green area indicates where our model outperforms the naive one, the red area indicates the opposite.

of people with insufficient food consumption up to 30 days into the future with higher accuracy than a naive approach solely based on the last measured prevalence, at least in places where enough training data are available to inform the model. The number of available historical observations proved to be a key element in forecasting success. Even for places with more than 10,000 available training points, which is not an extremely large number but still enough to provide reasonable results in other contexts⁶⁶, the phenomenon seems to be too complex for the algorithms to learn meaningful patterns.

Besides the external shocks related to local socio-economic conditions, it is important to note that all countries under study experienced the global effects of the COVID-19 pandemic since early 2020. The pandemic has significantly impacted food security on a global scale⁶⁷, disrupting supply chains, limiting access to food due to the adoption of non-pharmaceutical interventions, and increasing the need for food assistance⁶⁸. In our analysis, we did not include epidemiological variables to model the impact of the COVID-19 pandemic—such as reported cases or deaths—because those epidemiological indicators are likely to be unreliable in the countries under study, due to the lack of adequate surveillance systems⁶⁹. On the other hand, we expect the effects of the pandemic to be mainly captured by market price trends⁷⁰, which indeed markedly increased in all countries and in all regions as shown in Fig. S9 of the Supplementary Information. As SARS-CoV-2 continues spreading worldwide, the world economy still suffers from the consequences of the pandemic and its long term effects are hard to predict. Further research will be needed to investigate the complex interplay between the COVID-19 pandemic and food security.

Forecasting research within the humanitarian context has only recently started to attract attention from scholars⁷¹. In this context, our study represents an initial step towards the application of forecasting approaches to food insecurity at a high spatial and temporal granularity. Our results confirm that nowcasting or one-step-ahead forecasting are feasible, as reported in recent studies^{36,38}, but long-term forecasts are challenging and strongly conditioned by data availability.

The methods presented in this study come with limitations, and they could be further improved through several approaches. First, results could be compared with those obtained by other methods such as the autoregressive integrated moving average with exogenous inputs models (ARIMAX). Secondly, more complex forecasting methods could potentially lead to a greater forecasting accuracy, for instance through the use of deep learning techniques⁷². Additionally, hybrid methods, combining both statistical and ML features, could achieve a better forecasting performance⁶⁶. Finally, forecasting models could benefit from the inclusion of additional external predictors and in particular from the availability of novel data streams, such as mobile phone data⁷³ or the automated text mining of news⁷⁴.

Another important limitation of this study is the focus on one food insecurity indicator only, namely the prevalence of people with insufficient food consumption obtained from the FCS. As discussed above, a variety of indicators exist, each capturing different dimensions of food insecurity. Future work should therefore aim at developing forecasting models for other indicators too, when enough data is available. Additionally, the spatial resolution of this study was bounded due to data availability for first-level administrative units only. However, when second or third level administrative unit data becomes available, forecasting models at higher spatial resolutions should be developed. Finally, future studies should aim at going beyond the limitation of a 30-day forecasting horizon and propose methods that can forecast up to 2–3 months into the future.

In conclusion, our study presents a simple, yet fundamental message for governments and humanitarian organizations on the power of the data they collect: collecting data on a regular basis for long enough periods of time and across enough different geographic areas does not only make it possible to monitor the evolution of a situation in near real-time but also to inform forecasting models that would make it possible to produce estimates of how the situation is likely to evolve in the near future. This means that decision makers would have access in advance to information on areas most at risk of a deterioration in the food security situation, allowing for a more timely response. Predictions should of course be used with caution and considered only as an indication of what may happen in the near future, hence informing preparedness efforts by suggesting a need for further in-depth assessments of the food security situation.

Methods

Target indicator. The indicator whose time-evolution we aim to predict is the daily prevalence of people with insufficient food consumption in a given sub-national geographical area. This prevalence is obtained as the weighted share of households in the area that are found to have poor or borderline food consumption according to the Food Consumption Score (FCS)^{41,75}. The FCS is obtained through household surveys by asking how often, during the previous 7 days, a household has consumed food items from different food groups (main staples, pulses, vegetables, fruit, meat and fish, milk, sugar, oil and condiments). Consumption frequencies are then summed up in a weighted fashion, where each food group is weighted according to its nutritional level (with more nutritious foods having higher weights), resulting in the FCS. Thresholds are then applied to label each household as having poor, borderline or acceptable food consumption (as further detailed in Section 3.1 of the Supplementary Information), allowing to eventually compute the prevalence of people in a given area with insufficient (i.e. poor or borderline) food consumption.

The time series analyzed in this study were obtained by WFP through daily computer-assisted telephone interviewing (CATI) surveys. Informed consent was obtained from all interviewed subjects, all data collection protocols were approved by the World Food Programme's Hunger Monitoring Unit, and methods were carried out in accordance with its guidelines and regulations⁴⁸. Sample sizes were determined by WFP by taking into account modality and adhering to IPC guidelines for a good level of reliability (i.e. as close to 150 households per strata as possible)⁴⁰. They initially utilize Random-Digit Dialing (RDD) to obtain the most random selection of respondents as possible, while applying some filters to ensure the required geographic and sociodemographic

distributions (i.e. not all households reached are actually interviewed, only those matching the specific characteristics needed). This enables WFP, over time, to build a representative sample, and then transition to panel surveys after the initial months of implementation⁴⁸.

In order to compute a statistically representative prevalence of people with insufficient food consumption at sub-national and daily resolution, a rolling window approach is used. That is, for each geographical area, the prevalence of people with insufficient food consumption for a given day is obtained as the weighted share of households with poor or borderline food consumption interviewed during the previous d days, where d varies by country (values are reported in Supplementary Table S2). Missing values in the time series are inferred through linear interpolation. Post-stratification weights are applied by WFP to compute the final share of households with poor or borderline food consumption in a given area and time window, in order to mitigate sampling and modality bias, as detailed in⁴⁸. This is done through weighting of the data to account for the under-representation of certain demographics. Population weights are applied to compensate for administrative areas that are under- or over-sampled, while demographic weights are introduced to mitigate selection bias and compensate for under-represented households (e.g. low-income or less-educated households). The final weight is given by the product of the two, when both need to be applied. The present study only uses the final aggregated data, that is the estimated percentage of people in a given area with insufficient food consumption.

The available data covers six countries over the years 2018–2022. The temporal resolution is daily and the spatial resolution is that of first-level administrative units. The length of the time series and the number of geographical areas covered varies by country and is as follows: Burkina Faso (907 days, 13 areas), Cameroon (977 days, 10 areas), Mali (865 days, 7 areas), Nigeria (1140 days, 3 areas), Syria (1280 days, 12 areas) and Yemen (1340 days, 20 areas).

Permutation entropy. We employ the *Permutation Entropy* (PE) as a model-free measure of time series predictability³⁰. The main assumption of this approach is to measure the Shannon entropy through the probabilities of encountering trend patterns within the time series. For this reason, the PE first categorizes the continuous time series X in a small set of symbols or alphabet according to their trends. Let $x(i)$, $i = 1, \dots, N$, denote sequences of observations from a system X . For a given, but otherwise arbitrary i , m amplitude values $X_i = \{x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)\}$ are arranged in an ascending order where τ denotes the time delay, and m is the embedding dimension. Each X_i is then mapped onto one of the $m!$ possible permutations. The PE of the time series X is given by the Shannon entropy on the permutation orders:

$$H = - \sum_{\pi} p_{\pi} \log(p_{\pi}) \quad (1)$$

where p_{π} is the probability of encountering the pattern associated with permutation π . An important convenience of symbolic approaches is that they discount the relative magnitude of the time series⁷⁶. This is important in our case because different geographical units can differ largely in food insecurity prevalence. The embedding dimension m and the time delay τ are to be set in order to derive a reliable state space. There exist different procedural approaches in order to deal with this setting decision^{77,78}. In order to find the appropriate embedding dimension for clustering a set of time series, we follow the instructions proposed by Scarpino and Petri³⁰. The time delay is fixed to $\tau = 1$ in order to get results from continuous intervals. Finally, the metric used is the predictability defined as $\chi = 1 - H$. The closer to 1 the χ is, the more regular and more deterministic the time series is. Contrarily, the smaller χ is, the more noisy the time series is. As suggested by Scarpino and Petri³⁰, we analyzed the predictability as a function of the length of each time series. Focusing on the predictability over short timescales, we average H over temporal windows by selecting 1000 random points from each time series and calculating H for windows of length 10, 11, 12, ..., 100 days.

Independent variables. The following variables were defined to be considered as input features for the forecasting models. The same spatial and temporal coverage of the target indicator is used.

Prevalence of people using crisis or above crisis food-based coping. This prevalence is obtained as the weighted share of households in a given sub-national geographical area that are found to have a reduced Coping Strategy Index (rCSI) greater than or equal to 19^{60,75}. The rCSI is obtained through household surveys by asking if and how often, during the previous 7 days, a household had to adopt the following coping behaviors: relying on less preferred or less expensive food, borrowing food from relatives or friends, limiting portion sizes, restricting adults' consumption in order for small children to eat and reducing the number of meals eaten in a day. The rCSI is then obtained as a weighted sum of these frequencies, where weights are based on the severity of the strategy, as further detailed in Section 3.2 of the Supplementary Information. The survey data used to build this variable is the same as for the target indicator. A rolling window approach to compute a statistically representative prevalence of people using crisis or above crisis food-based coping at sub-national and daily resolution is also applied, and missing values are interpolated through linear regression. The same post-stratification weighting schemes to mitigate sampling and modality bias are also applied. The present study only uses the final aggregated data, that is the estimated percentage of people in a given area using crisis or above crisis food-based coping.

Conflict-related fatalities. The number of conflict-related fatalities in a given geographical area is obtained from the Armed Conflict Location and Event Data Project (ACLED), a publicly available near-global repository of reported conflict events and related fatalities⁶¹. Since each daily value of the target indicator is based on data collected during the previous d days, the number of fatalities associated with the same date and area is also obtained

by summing all fatalities reported in the same area during the same d days. Further details are reported in Section 3.3 of the Supplementary Information.

Market prices. Monthly prices of cereals and tubers are obtained from WFP's publicly available Economic Explorer (https://dataviz.vam.wfp.org/economic_explorer/prices). Cereal and tubers prices for each geographical area and date are obtained by averaging normalized prices (in local currency) across all markets within the area. Further details are reported in Section 3.4 of the Supplementary Information.

Weather variables. In order to measure the performance of the agricultural season, and more specifically whether the rainfall season is drier or wetter than average, and its impact on the vegetation status, for each geographical area and date, we consider the following weather variables, which are defined and computed by WFP as 10-day measurements, for each first-level administrative unit, and made publicly available through its Seasonal Explorer (https://dataviz.vam.wfp.org/seasonal_explorer/rainfall_vegetation/help): the amount of rainfall in mm, its 1-month and 3-month anomalies with respect to the historical average during the same period of the year (expressed in percentage), the normalized difference vegetation index (NDVI), and its anomaly (defined as for rainfall but considering 10-days only since effects of previous rainfall are already integrated by vegetation itself). Further details are reported in Section 3.5 of the Supplementary Information.

Religious observances. Ramadan is a religious observance celebrated by that the majority of the population in the analyzed countries during which food consumption increases. For each date and geographical area we therefore create a variable that takes into account the number of days, within the d days considered to obtain the prevalence of people with insufficient food consumption for the same date and area, that fall within the Ramadan observance period. This variable therefore spans between 1 and n during and after Ramadan, and is otherwise equal to zero during the rest of the year.

Population. The latest population estimate provided by WFP for each geographical area is used as a static variable.

Geographical area identifiers. The total area of each geographical unit, its latitude and longitude, and its waterways size, are also used as static variables.

Temporal identifiers. Temporal information (day, month and year) on the forecasting horizon is also included.

Preliminary feature selection. The proposed independent variables were defined based on expert opinion, and only a minimal preliminary selection was performed to avoid the presence of highly correlated variables within the same category (e.g., fatalities, market prices, weather, etc.). Since the only category that contains more than one variable is weather, this preliminary selection was performed only on the corresponding five variables: rainfall, 1-month rainfall anomaly, 3-month rainfall anomaly, NDVI and NDVI anomaly. We computed the Pearson's correlation coefficient r between each pair of the above five variables and for each pair where $|r| > 0.45$, one of the two variables was discarded, to avoid collinearity. Given the high information overlap among weather variables, the value 0.45 was chosen as an arbitrary conservative threshold to remove even moderate correlations.

As a result of this process, the rainfall 3-month anomaly was removed from all country-specific datasets but Nigeria's, the NDVI anomaly from Yemen's and Syria's, and the NDVI from the remaining four countries (see Supplementary Table S6), leaving rainfall, the 1-month rainfall anomaly and NDVI or its anomaly as the only remaining weather-related features for most countries, given the low level of correlation between them.

Finally, the Variance Inflation Factor (VIF) was computed for all remaining variables, including those in the other categories, to test the presence of multicollinearity. The resulting VIF values, reported in Supplementary Figures S15–S20, are all below 3, indicating no significant multicollinearity and hence allowing us to proceed with the obtained set of independent variables.

Forecasting. The core of this work revolves around the forecasting effort focusing on predicting the evolution up to 30 days into the future of our insufficient food consumption time series (No investigation into the forecast of the the independent indicators was performed because of the involvement of chain-of-events predictions (e.g. weather or market forecasts) and ethical issues around providing conflict predictions). To this aim, we chose to use the *eXtreme Gradient Boosting* (XGBoost) algorithm⁵⁹, a widely used state-of-the-art machine learning technique known for its high performance and flexibility. XGBoost belongs to the category of so-called ensemble learning approaches, which is a branch of machine learning methods that makes use of several models at once to produce a single better output. In the case of XGBoost, the base model is a decision tree, which is considered best-in-class for handling small to medium-sized data. A decision tree is a set of hierarchical choices which eventually lead to a final result, i.e. the prediction. Ensemble methods combine several decision trees to produce better predictive performance than utilizing a single decision tree. To create this collection of trees, XGBoost fits consecutive trees by, at every step, trying to solve for errors from the previous tree, using a gradient descent algorithm. The wider context of machine learning approaches used in the time series forecasting field and a more in-depth description of XGBoost can be found in Section 6 of the Supplementary Information.

The motivation behind the choice of this algorithm for this study is twofold. First, XGBoost can handle complex and non-linear relationships among the variables, which we expect to have in a complex phenomenon like

food insecurity. Secondly, XGBoost has a high degree of flexibility, which makes it the most suitable candidate for a prediction task that is meant to eventually run as an operational tool in near real-time. Specifically, XGBoost can handle missing values in the input variables, which is a feature that makes it possible to automatically run the algorithm on a daily basis, even when a few values might be missing because, for example, of delays in data availability.

Since XGBoost does not support a multi-output design, we developed 30 different models, one for each prediction horizon. For each date, the prediction framework is trained to predict levels of insufficient food consumption for a given day into the future based on the information available up to the date under consideration. For further details, see Section 6 of the Supplementary Information.

In order to implement our forecasting model based on the usual three stages of training, validation and testing, we adopt a k-fold cross-validation approach in a time-ordered fashion (i.e. the evaluation stage is applied to different historical periods).

The validation phase is implemented by splitting each of the n splits of training points of each sub-region into two time order preserving sets: the first 80% samples are used for training and the remaining ones for validation. Validation is performed independently across splits ensuring an unbiased approach. Our validation scheme aims to optimize the prediction framework by acting on two main configurations: model hyper-parameters and feature selection. The aim of this optimization is to find the configuration that returns the best performance as measured on a validation set. See Supplementary Table S8 for a detailed list of the explored hyper-parameters and values, and Supplementary Table S7 for the detailed list of independent variables and time lags considered. For further details, see Section 6 of the Supplementary Information.

Finally, in order to assess the goodness of the proposed forecasting model, its performance on the test sets is compared with a naive approach, where the predicted value at any given forecasting horizon is simply given by the last available value in the training and validation set, which represents the last available measured value before the start of the forecasting horizon.

Data availability

The data and code used to generate the results reported in this study are available in a public GitHub repository: <https://github.com/pietro-foini/ISI-WFP>. They are also available from corresponding author upon reasonable request.

Received: 16 September 2022; Accepted: 9 February 2023

Published online: 16 March 2023

References

- United Nations General Assembly. *Transforming Our World: The 2030 Agenda for Sustainable Development*. <https://sdgs.un.org/2030Agenda> (2015).
- FAO IFAD UNICEF WFP and WHO. *The State of Food Security and Nutrition in the World 2021. Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All*. <https://www.fao.org/publications/sofi/2021/en/> (2021).
- Food Security Information Network. *Global Report on Food Crises*. <https://www.fsimplatform.org/global-report-food-crises-2020> (2020).
- WFP Emergency Operations Division. *WFP Global Operational Response Plan: Update #3*. <https://www.wfp.org/publications/wfp-global-operational-response-plan-update-3-november-2021> (2021).
- Committee on World Food Security. *Global Strategic Framework for Food Security and Nutrition*. <https://www.fao.org/3/me498e/me498e.pdf> (2012).
- Clapp, J., Moseley, W. G., Burlingame, B. & Termine, P. The case for a six-dimensional food security framework. *Food Policy* **106**, 102164 (2021).
- Chavez, E., Conway, G., Ghil, M. & Sadler, M. An end-to-end assessment of extreme weather impacts on food security. *Nat. Clim. Change* **5**, 997–1001 (2015).
- Hasegawa, T. *et al.* Extreme climate events increase risk of global food insecurity and adaptation needs. *Nat. Food* **2**, 587–595 (2021).
- Balk, D. *et al.* Child hunger in the developing world: An analysis of environmental and social correlates. *Food Policy* **30**, 584–611 (2005).
- Grace, K., Davenport, F., Hanson, H., Funk, C. & Shukla, S. Linking climate change and health outcomes: Examining the relationship between temperature, precipitation and birth weight in Africa. *Glob. Environ. Change* **35**, 125–137 (2015).
- Shively, G. E. Infrastructure mitigates the sensitivity of child growth to local agriculture and rainfall in Nepal and Uganda. *Proc. Natl. Acad. Sci.* **114**, 903–908 (2017).
- Randell, H., Gray, C. & Grace, K. Stunted from the start: Early life weather conditions and child undernutrition in Ethiopia. *Soc. Sci. Med.* **261**, 113234 (2020).
- Dowd, C. *Conflict and Hunger* 1–8 (Springer International Publishing, Cham, 2020).
- Martin-Shields, C. P. & Stojetz, W. Food security and conflict: Empirical challenges and future opportunities for research and policy making on food security and conflict. *World Dev.* **119**, 150–164 (2019).
- Brück, T. & d'Errico, M. *Reprint of: Food Security and Violent Conflict: Introduction to the Special Issue* (2019).
- Drammeh, W., Hamid, N. A. & Rohana, A. Determinants of household food insecurity and its association with child malnutrition in sub-Saharan Africa: A review of the literature. *Curr. Res. Nutr. Food Sci. J.* **7**, 610–623 (2019).
- Gundersen, C. & Ziliak, J. P. Food insecurity and health outcomes. *Health Aff.* **34**, 1830–1839 (2015).
- Farrell, P., Thow, A. M., Abimbola, S., Faruqui, N. & Negin, J. How food insecurity could lead to obesity in lmics: When not enough is too much: A realist review of how food insecurity could lead to obesity in low-and middle-income countries. *Health Promot. Int.* **33**, 812–826 (2018).
- Shayo, F. K. & Lawala, P. S. Does food insecurity link to suicidal behaviors among in-school adolescents? Findings from the low-income country of sub-Saharan Africa. *BMC Psychiatry* **19**, 1–8 (2019).
- Falkendal, T. *et al.* Grain export restrictions during COVID-19 risk food insecurity in many low-and middle-income countries. *Nat. Food* **2**, 11–14 (2021).
- Mueller, V. *et al.* Food insecurity and COVID-19 risk in low- and middle-income countries. *Appl. Econ. Perspect. Policy* **44**, 92–109 (2022).

22. Rosegrant, M. W. & Cline, S. A. Global food security: Challenges and policies. *Science* **302**, 1917–1919 (2003).
23. WFP. *Food Security Analysis*. <https://www.wfp.org/food-security-analysis>. Accessed 22 June 2022.
24. WFP. *HungerMap LIVE*. <https://hungermap.wfp.org/>. Accessed 22 June 2022.
25. Elliott, G. & Timmermann, A. Forecasting in economics and finance. *Ann. Rev. Econ.* **8**, 81–110 (2016).
26. Timmermann, A. Forecasting methods in finance. *Annu. Rev. Financ. Econ.* **10**, 449–479 (2018).
27. Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: Case study of 2009 a/h1n1pdm. *BMC Med.* **10**, 1–31 (2012).
28. Kraemer, M. U. *et al.* Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat. Microbiol.* **4**, 854–863 (2019).
29. Perrotta, D., Tizzoni, M. & Paolotti, D. Using participatory web-based surveillance data to improve seasonal influenza forecasting in Italy. In *Proceedings of the 26th International Conference on World Wide Web* 303–310 (2017).
30. Scarpino, S. & Petri, G. On the predictability of infectious disease outbreaks. *Nat. Commun.* **10**, 898 (2019).
31. Wanner, N., Cafiero, C., Troubat, N. & Conforti, P. *Refinements to the FAO Methodology for Estimating the Prevalence of Under-nourishment Indicator* (FAO, 2014).
32. FAO IFAD UNICEF WFP and WHO. *The State of Food Security and Nutrition in the World 2020. Transforming Food Systems for Affordable Healthy Diets*. <https://www.fao.org/publications/sofi/2020/en/> (2020).
33. Mwebaze, E., Okori, W. & Quinn, J. A. Causal structure learning for famine prediction. In *2010 AAAI Spring Symposium Series* (2010).
34. Okori, W. & Obua, J. Machine learning classification technique for famine prediction. In *Proceedings of the World Congress on Engineering*, Vol. 2 4–9 (Citeseer, 2011).
35. Lentz, E., Michelson, H., Baylis, K. & Zhou, Y. A data-driven approach improves food insecurity crisis prediction. *World Dev.* **122**, 399–409 (2019).
36. Martini, G. *et al.* Machine learning can guide food security efforts when primary data are not available. *Nat. Food* **3**, 716–728 (2022).
37. Andree, B. P. J., Chamorro, A., Kraay, A., Spencer, P. & Wang, D. *Predicting Food Crises* (The World Bank, 2020).
38. Wang, D., Andree, B. P. J., Chamorro, A. F. & Girouard Spencer, P. *Stochastic Modeling of Food Insecurity* (The World Bank, 2020).
39. Westerveld, J. J. *et al.* Forecasting transitions in the state of food security with machine learning using transferable features. *Sci. Total Environ.* **786**, 147366 (2021).
40. Partners, I. G. *Integrated Food Security Phase Classification: Technical Manual Version 3.0: Evidence and Standards for Better Food Security Decisions* (Food and Agriculture Organization of the United Nations, 2019).
41. WFP. *Food Consumption Analysis. Calculation and Use of the Food Consumption Score in Food Security Analysis*. https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197216.pdf (2008).
42. Vaitla, B. *et al.* The measurement of household food security: Correlation and latent variable analysis of alternative indicators in a large multi-country dataset. *Food Policy* **68**, 193–205 (2017).
43. FAO. *The Food Insecurity Experience Scale*. <https://www.fao.org/in-action/voices-of-the-hungry/fies/en/>. Accessed 07 Dec 2022.
44. Backer, D. & Billing, T. Validating famine early warning systems network projections of food security in Africa, 2009–2020. *Glob. Food Secur.* **29**, 100510 (2021).
45. Bowling, A. Mode of questionnaire administration can have serious effects on data quality. *J. Public Health* **27**, 281–291 (2005).
46. Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P. & Zezza, A. High-frequency phone surveys on COVID-19: Good practices, open questions. *Food Policy* **105**, 102153 (2021).
47. Lamanna, C. *et al.* Strengths and limitations of computer assisted telephone interviews (CATI) for nutrition data collection in rural Kenya. *PLoS ONE* **14**, e0210050 (2019).
48. WFP. *The World Food Programme's Real-Time Monitoring Systems: Approaches and Methodologies*. <https://docs.wfp.org/api/documents/WFP-0000135070/download/>. Accessed 22 June 2022.
49. Brubaker, J., Kilic, T. & Wollburg, P. Representativeness of individual-level data in COVID-19 phone surveys: Findings from sub-Saharan Africa. *PLoS ONE* **16**, 1–27 (2021).
50. Glazerman, S., Rosenbaum, M., Sandino, R. & Shaughnessy, L. *Remote Surveying in a Pandemic: Handbook*. <https://poverty-action.org/sites/default/files/publications/IPA-Phone-Surveying-in-a-Pandemic-Handbook.pdf> (2020).
51. Ambel, A. A., McGee, K. R. & Tsegay, A. H. *Reducing Bias in Phone Survey Samples: Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries*. Policy Research Working Paper Series 9676 (The World Bank, 2021).
52. Bandt, C. & Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **88**, 174102 (2002).
53. Garland, J., James, R. & Bradley, E. Model-free quantification of time-series predictability. *Phys. Rev. E* **90**, 052910 (2014).
54. Zanin, M., Zunino, L., Rosso, O. A. & Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy* **14**, 1553–1577 (2012).
55. Kozak, J., Kania, K. & Juszczuk, P. Permutation entropy as a measure of information gain/loss in the different symbolic descriptions of financial data. *Entropy* **22**, 330 (2020).
56. Pennekamp, F. *et al.* The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecol. Monogr.* **89**, e01359 (2019).
57. Zunino, L. *et al.* Commodity predictability analysis with a permutation information theory approach. *Phys. A* **390**, 876–890 (2011).
58. Food Security Information Network. *Global Report on Food Crises*. <https://www.wfp.org/publications/global-report-food-crises-2021> (2021).
59. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (2016).
60. WFP. *The Coping Strategies Index: Field Methods Manual*. https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp211058.pdf (2008).
61. Raleigh, C., Linke, A., Hegre, H. & Karlsen, J. Introducing acled: An armed conflict location and event dataset: Special data feature. *J. Peace Res.* **47**, 651–660 (2010).
62. Liu, X. *et al.* Agricultural drought monitoring: Progress, challenges, and prospects. *J. Geogr. Sci.* **26**, 750–767 (2016).
63. D’Souza, A. & Jolliffe, D. Conflict, food price shocks, and food insecurity: The experience of afghan households. *Food Policy* **42**, 32–47 (2013).
64. Knippenberg, E., Jensen, N. & Constas, M. *Resilience, Shocks, and the Dynamics of Food Insecurity: Evidence from Malawi*. Technical Report, Working Paper (Cornell University, Ithaca, NY, 2018).
65. Headley, D. Rethinking the global food crisis: The role of trade shocks. *Food Policy* **36**, 136–146 (2011).
66. Makridakis, S., Spiliotis, E. & Assimakopoulos, V. The m4 competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **36**, 54–74 (2020).
67. Cardwell, R. & Ghazalian, P. L. COVID-19 and international food assistance: Policy proposals to keep food flowing. *World Dev.* **135**, 105059 (2020).
68. Workie, E., Mackolil, J., Nyika, J. & Ramadas, S. Deciphering the impact of COVID-19 pandemic on food security, agriculture, and livelihoods: A review of the evidence from developing countries. *Curr. Res. Environ. Sustain.* **2**, 100014 (2020).
69. Whittaker, C. *et al.* Under-reporting of deaths limits our understanding of true burden of COVID-19. *bmj* **375**, n2239 (2021).

70. Singh, S., Nourozi, S., Acharya, L. & Thapa, S. Estimating the potential effects of COVID-19 pandemic on food commodity prices and nutrition security in Nepal. *J. Nutr. Sci.* **9**, e51 (2020).
71. Altay, N. & Narayanan, A. Forecasting in humanitarian operations: Literature review and research needs. *Int. J. Forecast.* **38**, 1234–1244 (2020).
72. Lim, B. & Zohren, S. Time-series forecasting with deep learning: A survey. *Philos. Trans. R. Soc. A* **379**, 20200209 (2021).
73. Zufiria, P. J. *et al.* Identifying seasonal mobility profiles from anonymized and aggregated mobile phone data. Application in food security. *PLOS ONE* **13**, 0195714 (2018).
74. Kwak, H. & An, J. A first look at global news coverage of disasters by using the gdelt dataset. In *International Conference on Social Informatics* 300–308 (Springer, 2014).
75. The Integrated Food Security Phase Classification (IPC) Global Partners. *Technical Manual Version 3.0. Evidence and Standards for Better Food Security and Nutrition Decisions*. https://www.ipcinfo.org/fileadmin/user_upload/ipcinfo/docs/IPC_Technical_Manual_3_Final.pdf (2019).
76. Borge-Holthoefer, J. *et al.* The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Sci. Adv.* **2**, e1501158 (2015).
77. Cao, L. Practical method for determining the minimum embedding dimension of a scalar time series. *Phys. D* **110**, 43–50 (1997).
78. Kennel, M. B., Brown, R. & Abarbanel, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411 (1992).
79. Jordahl, K. *et al.* geopandas/geopandas: v0.9.0. <https://doi.org/10.5281/zenodo.4569086> (2021).

Acknowledgements

PF, DP and MT gratefully acknowledge the Lagrange Project of the ISI Foundation funded by the CRT Foundation. EO and GM would like to acknowledge her WFP colleagues for the fruitful discussions.

Author contributions

E.O., D.P. and M.T. supervised the study. P.F. and G.M. analyzed the data. P.F. developed the code and ran the models. All authors analysed the results and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29700-y>.

Correspondence and requests for materials should be addressed to E.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023