

Response to Reviewers

We thank the reviewer for the detailed and thoughtful feedback, which forced us to engaged more with issues of spatial scale, household vs population food security, as well as a more broad body of literature thn we originally examined. We believe that the paper is now much improved and will be of interest to a wider body of researchers, with more clear explanations of the methodology and spatial scale, and more detailed discussions of issues like population, urbanization, and climate.

This paper presents a global assessment of food security outcomes and prediction for the years to come. It leverages a useful household-level food security dataset and gap fills with a random forest-model to predict food security outcomes in missing locations. While it has potential to add important findings to the literature, there are several methodological gaps that need to be addressed before it can be considered for publication. The General Comments below reflect broader specific locations in the text where I take issue. That said, I have not fully addressed line-by-line critiques yet because the broader concerns are too numerous. I highly encourage the authors to not be discouraged, as I see this work as immensely valuable, but I want them to produce the most robust science possible.

First, there is a growing body of literature that shows wide variation estimates of population across gridded gridded population products (e.g., Chen et al. 2020; Bustos et al. 2020; Tuholske et al. 2019) and recently published fitness-for-use guidelines (see Leyk et al. 2019). The authors do not clearly explain why they chose WorldPop (as opposed to Landsat, ESRI's World Population Estimates, the Global Human Settlement Layer, High Resolution - see www.popgrid.org for more). Furthermore, it is not clear how the Worldpop data is even used. It is mentioned in the Main Text, but I do not see it in your Supplemental Materials. Population estimates should be re-run using other gridded population datasets as I suspect that they will yield markedly different results.

We regret that we did not make it more clear how the population data was used - we have now made that more clear in the text, and that may have also made it more apparent why we chose the WorldPop dataset. We use gridded population data to estimate demographic characteristics for administrative areas, and chose to use WorldPop

because it has data on age and sex. Examining the other datasets you mention and searching through that very useful review by Leyk et al., it seems that GPWv4 also includes such demographic data, but more coarsely aggregated (age categories are across wide ranges, such as 15-49, whereas WorldPop bins ages every five years). Moreover, WorldPop is an annual product that we can match to the specific year of our survey data, whereas GPWv4 is only available for 2015. We have now made it more apparent in the *Disaggregation* subsection of the *Methods* section why we chose WorldPop, and have added some discussion and relevant citations from your review.

Additionally, we also used population data at the administrative area level to estimate the food insecure population totals after modeling rates of food insecurity. This data came from the [Subnational Human Development Database \(SHDI\)](#), which in turn is derived from national censuses. Because this analysis is conducted at the admin-area level, we believe it is not necessary to use high-resolution gridded data to estimate population totals and the SHDI data is sufficient.

Second, like the population data employed, I have strong concerns with the temperature and precipitation data and methodology employed here. Why use TerraClimate over the widely used and higher resolution CHIRPS (Funk et al. 2015) or other climate reanalysis (ERA5, MERRA-2, etc.) and the new CHIRTS-Max (Funk et al 2019) or CHIRTS-daily (Verdin et al. 2020) high-resolution temperature products to establish your baseline? What temporal resolution is ‘mean temperature’? Surely it is widely established the annualized climate data over large distance data does not reflect the true nature of coupling between household food security outcomes and climate (see Shukla et al. 2020 for one recent example). I point to the example of Sternberg 2012 that clearly shows how climate shocks/variability across the planet can telescope to food security outcomes at great distances. How climate (across spatial and temporal scales) affect local price and labor is key here, yet you do not even discuss price at all, much less employ it, in your modeling paradigm. Yet we know that specifically for urban areas, where half the planet lives, food security outcomes are highly dependent on labor and price (see Blekking et al. 2020 for an example). Also, what about seasonality (again see Shukla et al. 2020 for one example)? None of these crucial aspects of household-level food security outcomes are clearly addressed in the current version of the manuscript. I suspect that the reason your ‘climate’ variables do not predict your outcome is because you spatially and temporally smooth your data to the point where it will no matter what not capture the true nature of how climate shocks impact food security (e.g. huge spatiotemporal coupling between where food is grown, when it is grown, how much the price is dictated by local, national, and global market forces, etc.). See Niles et al. 2021 for a much more robust approach.

We thank you for this detailed feedback, and we agree that the approach to modeling the impacts of climate on food security is too

simple to capture all of the pathways by which the climate affects food security. However, we are in many ways constrained by our survey data, which is only available at the annual temporal scale and only reported at the national spatial scale. This makes it much harder to trace the impacts of climate on food security and necessitates the smoothing of our climate data. The papers you cite by Niles, Blekking, and Shukla all seem to have indicators of food security available at fine spatial scales (Niles uses DHS clusters/community GPS points, Blekking uses household GPS points) and fine temporal scales (Niles uses monthly data, Shukla uses Dekads).

We also note that it is important to consider the scope of our analysis. Weather shocks as well as labor and food prices certainly explain a lot of the variation in food security between households and across seasons. However, the scope of this analysis is to explore drivers of food insecurity across the globe and over decadal timeframes. We are especially interested in modeling future trends in food security at these broad spatial and temporal scales. We therefore use GCM projections, which are useful insofar as they are indicative of shifting baselines, but are unable to anticipate actual droughts or floods that would impact food insecurity in 2030.

Given these considerations of how the scale of our data (annual, national sampling) and scope of our analysis (modeling food insecurity globally and out to 2030) makes it infeasible to explore the effect of climate on food insecurity beyond shifting baselines, we do not think that using other datasets like ERA-5 or MERRA-2 would add much value. (Also, the CHIRPS dataset, while excellent, is not truly global and unavailable for certain latitudes, so couldnt be applied to this global analysis). However, you raise very important considerations in your feedback and we therefore make two improvements to the paper: (1) in addition to the survey year's rainfall totals and precipitation averages, we add a feature for the previous year's total rainfall and mean temperature, and additionally (2) we add a paragraph to the discussion highlighting many of the points you raise and further outlining the considerations that led to the analytical approach we took.

Third, Random Forrest models suffer from training data bias. If the underlying training data is flawed, the Random Forest Model has no way of identifying this. What quality control is used to determine if your survey data is correct? One way to address this is to compare the Gallup data with other sources like DHS or LSMS surveys from the same year. Do they both measure similar national-level food security outcomes? I am also curious to know how your models preform with simple linear or logistic regression? Do you get the same results? I recognize Random Forest are nonparametric, but I would be keen to see how linear models preform with controls. Also, what happens when you trade out wasting/stunting

as your predicted outcome versus the FIES-based metric?

Unfortunately, the Gallup World Poll (GWP) is currently the only source of nationally-representative data on the FIES, as this metric is not collected by surveys like the DHS and LSMS (although as the FIES has been adopted as an SDG indicator, this is changing, see [here](#)). Thus, we cannot use those datasets to validate our Random Forest model. However, we note that we do take care to conduct cross validation at the country level. We leave out all surveys a given country and then evaluate the model's performance on that country, and find our model it still performs favorably.

Additionally, we have updated the paper to now model the FIES using LASSO regression (in the supplement), and note that, compared to linear regression, the random forest model performs quite well. We also now add some discussion around the importance of accounting for interactions between covariates based on a comparison between random forest and linear regression.

Fourth, the spatial nature of your modeling is confusing and not clear. Are metrics averaged at the sub-national scale (if so what level of analysis and for which country)? How is degree of urbanization quantified (e.g. are Worldpop pixels masked at some threshold)? Is the Gallup representative at the national level? District-level? What controls are used to ensure that the Gallup data is a representative sample (Not even explained in the Supplemental Materials)? Your methodology (including how you address issues of spatial scale) needs to be more clearly described in the Main Text and the Supplemental Material before this paper can be considered for publication.

To add more clarity, we have changed the phrasing in most places from “subnational” to “administrative area”, as this might improve some of the confusion around the spatial scale of our analysis. We have also added a paragraph to the methods section giving a high-level overview of the analysis, clearly stating the scale of the study (year x admin area). The Gallup data is designed to be representative at the national level through a weighting schema based on demographic characteristics, and we used the same methodology to get estimates at the administrative area level. Gallup records whether individual respondents are in urban or rural areas based on the individuals own reporting. We then use global gridded data from Jones and O'Neill (2016) published in *Environmental Research Letters* to estimate the percent of the population in urban areas for each administrative area through the year 2030.

Fifth, I take issue with this being the first global food security paper at a sub-national level (Pg 12, ln 11). Niles et al. is another similar paper, that is near global, with a much more robust modeling framework. Additionally, FEWS-net and WFP provide these assessments regularly or real time. Specifically, you do

not present any sub-national findings. What are they? Why is sub-national important and how do your results advance our understanding of food security over space and time? Surely there may be important nuances in the sub-national data (e.g. Does Nairobi look like Dhaka? Is your assumption that two households with similar characteristics, irrespective of geography, correct? Do your results show that this assumption is correct?).

When stating that this is the first global picture of food insecurity, we meant that this was the first analysis using the FIES metric, and have ensured that anywhere we make that claim through the paper, we include that important caveat. We also believe that our use of the word “global” indicates that our analysis has estimated food insecurity in all countries or nearly all countries, which was not done by Niles et al (who only used 19 countries) and is not done by the FEWSNET reports (which only targets certain developing regions). Thus, we believe that this claim is valid, and we believe future readers would interpret the word “global” the same way. However, if you disagree, please let us know in a future review and we can change the language. We have also added more description of the subnational findings to the results section.

Finally, and importantly, how are you defining ‘urban’? (See Cattaneo et al. 2021 for further details). This is not clear in your Supplementary Material or main text. You do not address the growing body of literature that clearly demonstrates important differences between ‘rural’ and ‘urban’ food security measurement and outcomes (see Blekking et al. 2020, Tuholske et al. 2020, Haysom & Tawodzera 2018, Cockx et al. 2018) that are extremely important for stakeholders and policy development (Battersby 2017) and poverty alleviation (Christiaensen et al. 2017).

Our data on urban and rural areas comes from Jones and O'Neill 2016, which use as their data source the Global Rural-Urban Mapping Project (GRUMP), v1. GRUMP in turn bases their definition of urban on night-time lights and buffered settlement centroids. We have now made that more explicit in the text. We note the importance of urbanization in determining food security outcomes and now discuss how an analysis at this the admin area scale cannot capture all of the impacts of finer processes like urbanization and climate. We have also added some of the literature you've highlighted to the paper.

A few specifics:

Pg. 3, Ln 35 - 38: This assumption needs strong justification. Recent urban food security studies show wide heterogeneity in food security outcomes among ‘low-income’ households depending on the food security metric used (see Blekking et al. 2020, Tuholske et al. 2020, etc.). ‘Urbanization level’ at an administrative-unit level (if that is what you are using...it's not clear), thus, may not be good

predictor of food security outcomes.

We agree that there is a substantial variation in food security status between urban and rural areas, even low income ones. However, we also include data on other available demographic factors, including gender and age, to perform our disaggregation. Moreover, we are not analyzing individual households - we are just making population-level estimates of rates of food insecurity at the administrative area scale, based on samples that include thousands of households. Some assumptions are built into this process, but they are common to all survey weighting methods, as we now note in the text.

Pg. 4, Ln 20 - 23: Please expand your justification as to why you selected these climate datasets. Why is spatial averaging over large length scales appropriate to predict food security outcomes? Why is annual average precipitation and (I assume annual) average temperature appropriate?

In addressing your previous higher-level comment, we have added significantly more discussion around climate and spatial scale to the paper. We add a note to this section highlighting that, given the scale of our food security data and our analysis, we are mostly focusing on climatic baselines as they relate to agro-ecological conditions

Pg. 7, Ln 19 - 30: This results is in contrast to FEWS NET outlook that the those at risk to famine rose by 85 million people from 2015 - 2019 (I can't find the brief, but here is an article with the numbers: <https://www.sciencemag.org/news/2020/04/how-team-scientists-studying-drought-helped-build-world-s-leading-famine-prediction>). This was before covid. Why do your numbers differ?

This is an interesting point. We note that the paragraph you take issue with is specifically focusing on the period of 2010 to 2030, and our figures, which include the range 2015-2019, show increases in severe and moderate food insecurity over this period, jibing with CHG's findings. We also took some effort to compare our results to other reports of trends in food security, focusing mostly on the metric of undernourishment in the FAO's State of Food Insecurity report. We now have a reference to the science news article you mentioned.

Pg. 9, 3.2: You have sub-national data, and you make assumptions that the characteristics of two similar households have the same food security outcomes, yet you present your findings at national level. An increasing body of research (e.g. ref) is moving away from national-level estimates as they mask the true nature of food security outcomes and are not relevant for policy maker or targeted interventions. Please justify these findings given the high degree of spatiotemporal heterogeneity in food security outcomes.

First, we share some of your concerns about the inadequacy of national-level food security estimates, which is precisely why we

chose to disaggregate our analysis to the admin area level. We further note that our data was originally available at the national scale, and we used demographic data to disaggregate it (See section 2.1). Disaggregating to the admin 1 area is possible with our data, but a household or individual-level analysis really is not.

We also disagree that “national-level estimates... are not relevant for policy makers”. While local practitioners focus on differences in food security between individuals or households, many multinational practitioners pay close attention to country-level (or admin-level) estimates of food security. We note the FAO’s State of Food Insecurity (SOFI) reports, which are widely read, as well as the fact that the SDG targets, which are key metrics for policymakers and which include the FIES for SDG2, are monitored at the national level (see: <https://sdg-tracker.org/zero-hunger>). We discuss both the SOFI estimates and the relevance to SDG2 throughout the paper. Moreover, as we mention in the acknowledgments, this project began out of a specific demand from an international policymaking organization. So we feel that our analysis meets a clear stated need by policymakers, and is conducted at the appropriate spatial scale given the data that we have.

References

- Battersby, J. (2017). MDGs to SDGs-new goals, same gaps: the continued absence of urban food security in the post-2015 global development agenda. *African Geographical Review*, 36(1), 115-129.
- Blekking, J., Waldman, K., Tuholske, C., & Evans, T. (2020). Formal/informal employment and urban food security in Sub-Saharan Africa. *Applied Geography*, 114, 102131.
- Bustos, M. F. A., Hall, O., Nedomysl, T., & Ernstson, U. (2020). A pixel level evaluation of five multitemporal global gridded population datasets: a case study in sweden, 1990-2015. *Population and environment*, 1-23.
- Cattaneo, A., Nelson, A., & McMenomy, T. (2021). Global mapping of urban-rural catchment areas reveals unequal access to services. *Proceedings of the National Academy of Sciences*, 118(2).
- Chen R, Yan H, Liu F, Du W, Yang Y. Multiple Global Population Datasets: Differences and Spatial Distribution Characteristics. *ISPRS International Journal of Geo-Information*. 2020;9: 637.
- Christiaensen, L., & Kanbur, R. (2017). Secondary towns and poverty reduction: refocusing the urbanization agenda. *Annual Review of Resource Economics*, 9, 405-419.
- Cockx, L., Colen, L., & De Weerd, J. (2018). From corn to popcorn? Urbanization and dietary change: evidence from rural-urban migrants in Tanzania.

World Development, 110, 140-159.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1), 1-21.

Funk, C., Peterson, P., Peterson, S., Shukla, S., Davenport, F., Michaelsen, J., ... & Mata, N. (2019). A high-resolution 1983-2016 T max climate data record based on infrared temperatures and stations by the climate Hazard center. *Journal of Climate*, 32(17), 5639-5658. Haysom, G., & Tawodzera, G. (2018). “Measurement drives diagnosis and response”: Gaps in transferring food security assessment to the urban scale. *Food Policy*, 74, 117-125.

Leyk, S., Gaughan, A. E., Adamo, S. B., Sherbinin, A. D., Balk, D., Freire, S., ... & Pesaresi, M. (2019). The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3), 1385-1409.

Niles, M. T., Emery, B. F., Wiltshire, S., Brown, M. E., Fisher, B., & Ricketts, T. H. (2021). Climate impacts associated with reduced diet diversity in children across nineteen countries. *Environmental Research Letters*, 16(1), 015010.

Shukla, S., Husak, G., Turner, W., Davenport, F., Funk, C., Harrison, L., & Krell, N. (2021). A slow rainy season onset is a reliable harbinger of drought in most food insecure regions in Sub-Saharan Africa. *Plos one*, 16(1), e0242883.

Sternberg, T. (2012). Chinese drought, bread and the Arab Spring. *Applied Geography*, 34, 519-524.

Tuholske, C., Caylor, K., Evans, T., & Avery, R. (2019). Variability in urban population distributions across Africa. *Environmental Research Letters*, 14(8), 085009.

Tuholske, C., Andam, K., Blekking, J., Evans, T., & Caylor, K. (2020). Comparing measures of urban food security in Accra, Ghana. *Food Security*, 1-15.

Verdin, A., Funk, C., Peterson, P., Landsfeld, M., Tuholske, C., & Grace, K. (2020). Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Scientific Data*, 7(1), 1-14.