Response to Reviewers

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 numerousR 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 Landscan, 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 out of all of the datasets you mention and that were surveyed by Leyk et al., WorldPop has the richest data on age and sex. 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

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, which derives from national censuses

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.

Third, Random Forrest models suffer from training data bias. If the underlying training data if 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?

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.

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?).

Finally, and importantly, how are you defining 'urban'? (See Catteneo 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).

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.

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?

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 cofid. Why do your numbers differ?

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.

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., Niedomysl, 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 Weerdt, 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.