Methods

The methodology for estimating food insecurity at the subnational level involved three steps:

1. Use FAO survey microdata to estimate subnational levels of food insecurity for all countries in the dataset as training data.
2. Estimate for factors related to wealth, the environment, health, and market access as covariates for all subnational areas from 2010 to 2030.
3. Use a random forest model to estimate global rates of food insecurity, subnationally, from the years 2010 to 2030.

# Disaggregating Individual-Level FIES Microdata

We started with 331 surveys from 77 countries from the years 2014 to 2018 from the [FAO microdata catalogue](https://microdata.fao.org/index.php/catalog/Food-Security). This survey data comes in the form of individual-level data, with approximately 1,000 individuals interviewed per survey. For each individual, data on their age, gender, education, locality (urban/rural) and national wealth quintile is collected, as well as whether they exhibited 8 behaviors indicative of food insecurity over the previous year. Based on these 8 behaviors, using methods [developed by the FAO](https://doi.org/10.1016/j.measurement.2017.10.065), the FAO estimates the probability that an individual is over the threshold for severe and moderate-to-severe food insecurity according to the Food Insecurity Experience Scale (FIES). To get estimates of national prevalences, the FAO assigns each individual a weight based on how their demographic characteristics are more or less representative of national characteristics compared to the sample of individuals interviewed, and then takes the weighted mean probability of being over thresholds for severe and moderate-to-severe food insecurity.

We use a similar methodology to the FAO, except we use subnational data on wealth, gender, age, and urbanization to calculate separate weights at each subnational level. We use each individual in the country sample for each subnational calculation with different weights each time. To estimate subnational prevalences of age and gender, we use data from [WorldPop](https://worldpop.org), for subnational prevalences of urban and rural population, we use data from [Jones & O’Neill (2016) *Environ. Res. Lett.*](https://doi.org/10.1088/1748-9326/11/8/084003), and for the subnational prevalences of wealth quintiles, we use data from [Demographic and Health Surveys](https://dhsprogram.com/), where available. This methodology allows us to estimate and model the FIES at a subnational level, and rests on the mild assumption that individuals of the same age, gender, urban/rural context and wealth quintile will have the same risk of food insecurity, no matter where in the country they are observed.

# Covariates

We use a dozen covariates characterizing food systems to model food insecurity rates over space and over time. For each covariate that is not time-invariant, we aim for data sources that have been peer-reviewed and include projections to at least the year 2030. For projections that come from scenario or simulation based methodologies, we use the most middle-of-the-road estimates. Thus, for climate projections, we use estimates from the Representative Concentration Pathway (RCP) 6.5, and for projections from the Shared Socio-Economic Pathways (SSP) framework, we use projections from SSP 2. An overview of the data sources is given in the table below, and an overview of different methods used for harmonizing and forecasting individual covariates are given below that.

* **Urban Percentage** - [Jones & O’Neill (2016) *Environmental Research Letters*](https://doi.org/10.1088/1748-9326/11/8/084003)
* **Stunting** - [Local Burden of Disease (2020) *Nature*](https://doi.org/10.1038/s41586-019-1878-8)
* **Wasting** - [Local Burden of Disease (2020) *Nature*](https://doi.org/10.1038/s41586-019-1878-8)
* **Mean Years of Schooling** - [Smits and Permanyer (2019) *Scientific Data*](https://doi.org/10.1038/sdata.2019.38), [KC (2017) *Global Environmental Change*](https://doi.org/10.1016/j.gloenvcha.2014.06.004)
* **GDP Per Capita** - [Smits and Permanyer (2019) *Scientific Data*](https://doi.org/10.1038/sdata.2019.38), [Dellink et al. (2017), *Global Environmental Change*](https://doi.org/10.1016/j.gloenvcha.2015.06.004)
* **Gini Coefficient** - [Rao et al. (2019) *Futures*](https://doi.org/10.1016/j.futures.2018.07.001)
* **Poverty Headcount Index** - [Cuaresma et al. (2018) *Palgrave Communications*](https://doi.org/10.1057/s41599-018-0083-y)
* **Water Scarcity** - [Greve et al. (2018) *Nature Sustainability*](https://doi.org/10.1038/s41893-018-0134-9)
* **Topographic Ruggedness** - [USGS (1996)](https://doi.org/10.5066/F7DF6PQS)
* **Mean Temperature** - [Abatzoglou et al. (2018) *Scientific Data*](https://doi.org/10.1038/sdata.2017.191), [Warszawski et al. (2014) *PNAS*](https://doi.org/10.1073/pnas.1312330110)
* **Mean Annual Precipitation** - [Abatzoglou et al. (2018) *Scientific Data*](https://doi.org/10.1038/sdata.2017.191), [Warszawski et al. (2014) *PNAS*](https://doi.org/10.1073/pnas.1312330110)
* **Malaria *P. falciparum* Mortality Rate** - [Weiss et al. (2019) *The Lancet*](https://doi.org/10.1016/S0140-6736(19)31097-9)

## Combining, Harmonizing and Extrapolating

For urban percentage, Gini coefficient, poverty headcount index, and water scarcity the dataset covered all of the years of the study.

For stunting, wasting, and malaria mortality rate, we estimate the Annualized Rate of Change (AROC) for each subnational area. The involves taking the rate of change between each pair of years in the dataset, and then taking the mean rate of change over the period for which data is available, giving greater weight to more recent years. We then estimate future changes in stunting, wasting, and malaria mortality, based on this rate of change.

For mean years of schooling, GDP per capita, as well as for projections of population, we had subnational historic data and future projections at the national level. We used observed trajectories in the historical inequalities among subnational areas within a country to estimate the future distribution of GDP, population, and schooling among subnational areas and disaggregate national-level future projections.

For temperature and precipitation, we combined historical observations with an ensemble of bias-corrected projections for the future climate from the Combined Model Intercomparison Project 5 (CMIP5).

## Accounting for COVID-19

Most of these datasets were created before the COVID-19 pandemic, and therefore their projections need to be adjusted in light of the impacts of the pandemic on the global economy and food system. The estimates of poverty headcounts and GDP per capita were revised based on World Bank estimates of the impact of the pandemic on country-level GDP growth for the years 2020 and 2021, with economic growth resuming its previously estimated rate of change after that time. Additionally, the Lancet has published estimates of the impact of the pandemic on child anthropometry, which we used to adjust estimates of stunting and wasting for 2020 and 2021, with those two variables return to their projected trajectories by 2022.

## Missing Data

In cases where countries were missing data on covariates, we used estimates from the most similar plausible country. With this is a rough method, but it permits us to make more reasonable estimates of global totals of the population that is food insecure than we could if we excluded the entire population of data-poor countries.

* For Syria, we used the same poverty headcount rates as Yemen, another conflict-affected middle eastern nation.
* For South Sudan, we used the same projections of GDP per capita and Gini coefficient as the Central African Republic, another landlocked and under-developed central African nation.
* For North Korea, we estimated the same level of education and GDP per capita as Tajikistan, another economically isolated, authoritarian Asian nation.

# Modeling Food Insecurity Over Time and Space

We trained random forest models on the subnational estimates of food insecurity prevalence, with separate models for severe and moderate-to-severe food insecurity. We estimated model hyperparameters using the settings that performed best with cross-validation. Overall, or model performed very well. With 10-fold cross validation, the model has an r2 of about 0.98, and on the full dataset, it has an r2 of greater than 0.999. Finally, we tested the model on national estimates reported by the FAO. These estimates do not include individual level-data and therefore were not included in model fitting at all. On this out-of-sample dataset, the model had an r2 of about 0.96.

# Data, Code, and Future Publications.

Add code for this analysis is available on [github](https://github.com/mcooper/wdl-fies).

Data can be accessed by contacting the [World Data Lab](https://worlddata.io/contact).

This analysis is currently under peer review and this page will be updated when it is published.