A Global Picture ofHunger

Matthew Cooper

Benjamin Müller

Carlo Cafiero

Juan Carlos Laso Bayas

Homi Kharas

Jesús Crespo Cuaresma

Using microdata from 75 countries collected by the FAO and Gallup World Poll for the Voices of the Hungry project, we model and provide predictions of levels of food security at the subnational level from 2010 to 2030. This is the first global assessment at a subnational level of the Food Insecurity Experience Scale, the food security metric most indicative of the lived experience of hunger. We find significant heterogeneity in food security around the world, ranging from less than 1% of the population being food insecure in high-income countries to parts of the developing world where as much as 3 out of 4 people are severely food insecure. Examining global temporal trends and accounting for the effects of the COVID-19 pandemic, we find that food insecurity has increased slightly over the past several years, but under middle-of-the-road assumptions for development and population change, food insecurity is expected to decline throughout the 2020s. This global decrease is largely driven by trends in South and East Asia, while some other parts of the world, particularly sub-Saharan Africa, are projected to experience increases in the prevalence of food insecurity throughout the next decade. Our model forecast of decreasing food insecurity through the 2020s contrasts with predictions from the FAO for other metrics of hunger.

# Introduction

Food security is a critical component of human flourishing and its importance as a global policy objective is reflected in the second Sustainable Development Goals (SDG 2, “Zero Hunger"). One of the indicators to track progress on the first target of this goal, to ensure access by everyone to safe, nutritious and sufficient food (SDG2.1), is the prevalence of moderate to severe food insecurity in the population based on the Food Insecurity Experience Scale, or FIES. The FIES was developed by FAO in 2013 as a global extension of pioneering work done on Latin America to develop a metric of food insecurity based on the perspective of people who experience hunger and conditions that produce hunger. Other indicators of food insecurity, such as FAO’s most commonly used estimate of undernourishment, are based on macro models of the mean and distribution of calories consumed in a population. These do not directly capture the actual lived exposure of individuals to food shortages.

The FIES is based on responses to polls and the data are limited to 75 countries where these have been conducted and microdata has been vetted and released by the FAO. The current practice is to assign the average regional prevalence rate for countries without survey data in order to compile regional and global aggregates. In addition, since the FIES is a relatively new metric, it does not have a sufficiently long time series to assess time trends with a reasonable degree of precision. To understand whether SDG 2 is being met, estimates of food insecurity outcomes in 2030 are required.

This paper contributes to shifting the research frontier by filling these two gaps in terms of data coverage and forecasts by developing a statistical model of FIES outcomes based on covariates that are available for territories where actual polling data are not available. Such a model can be employed to obtain forecasts to 2030 and assess likely future trends in food insecurity. We take advantage of the within-country spatial distribution of FIES reports to push the analysis beyond country level to a sub-national (Admin-1) level. To our knowledge, this is the first effort to model the FIES outcomes at that scale.

Food security has traditionally been difficult to measure, and this has led to an incomplete or inaccurate assessment of global hunger. Metrics of macro-health, such as anthropometric measures and mortality rates are correlated broadly with food insecurity and have been used for many years to monitor human well-being . However, these metrics are affected by other determinants of health such as the occurrence of infectious diseases, and are not meaningful at the scale of individuals or households. Other proxies for food security, such as food availability estimated from crop yields, are also inadequate because they only make rough estimates of food accessibility, and food insecurity can certainly occur in the absence of food availability decline . Moreover, these metrics are very sensitive to incorrect estimates of crop yields and food reserves at a national scale. Thus, global estimates of hunger and food insecurity based on these metrics carry forward similar flaws.

As researchers began to focus on food insecurity at the individual and household level, household microdata collecting information on household finances and consumption became a common proxy for food security . However, these efforts were criticized for being onerous, insufficiently comparable, as well as for ignoring subjective and experiential aspects of food security . This led to the emergence of several indicators designed to be rapidly deployable, and based on the lived experience of food security . These metrics include the Household Food Security Survey Module (HFSSM) , originally developed for use in the US; the Latin American and Caribbean Food Security Scale (ELCSA); and the Household Food Insecurity and Access Scale (HFIAS) .

Drawing on the insights derived in designing and implementing these novel food security metrics, the FIES was developed by the Food and Agricultural Organization (FAO) of the UN and is now recognized as a rapidly deployable and cross-culturally valid tool for understanding individual and household-level food insecurity . The FIES is based on a survey of eight behaviors indicative of food insecurity and hunger over the previous year, such as skipping meals or worrying about having enough to eat. Using a Rasch model , each individual in the survey is given a severity score that maximizes the likelihood of observing the responses given to the 8 questions asked in the survey. Once individuals are ranked in this fashion, thresholds for moderate and severe food insecurity are set that describe people’s experiences, and the prevalence of such food insecurity events can be computed. A moderately food insecure adult is likely to have compromised on food quality and variety as a result of lack of income, has been unsure about their ability to obtain food and has skipped meals or run out of food occasionally. A severely food insecure adult is likely to have run out of food and has gone an entire day without eating at times during the year. Because the model estimates the population over a threshold on a univariate food insecurity scale, the percentage of people over the threshold for moderate food insecurity includes people that are also over the threshold for severe food insecurity.

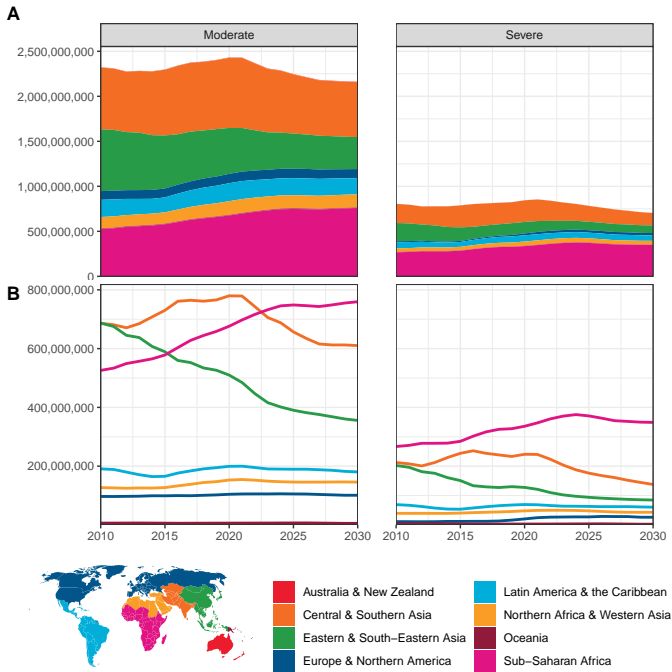
Since 2014, in partnership with Gallup World Poll, the FAO administered the FIES in surveys around the world. The FAO uses this data to estimate the prevalence of food insecurity, and reports national-level estimates of the percentage of the population over the thresholds for moderate and severe food insecurity. Drawing on individual-level data from these microdatasets in 75 countries, we use machine learning methods to estimate levels of food insecurity at the subnational level. The model is then used to forecast food insecurity to the year 2030 under a benchmark scenario.

# Results

## Food Insecurity over Time

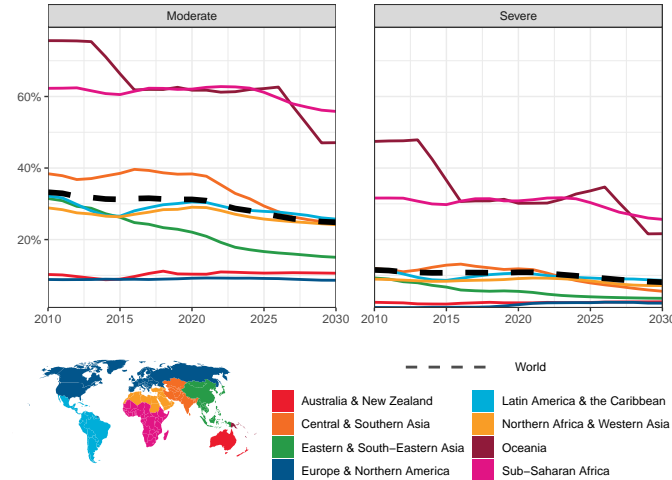
Globally, in the year 2020, we estimate that 829 million people are above the threshold for severe food insecurity, and an additional 1.5 billion people are above the threshold for moderate food insecurity (See Fig. [1](#fig:timeseries)). These numbers are higher than the estimates in the 2019 FAO State of Food Insecurity report, which reports about 700 million severely food insecure people and an additional 1.31 billion people experiencing at least moderate food insecurity in 2018.

While the total number of people experiencing moderate and severe food insecurity was roughly constant for the 2010s and even increased in latter part of that decade, our model predictions indicate that this figure will decline throughout the 2020s. However, this overall decline will not be uniform across the world. Our model projections show that moderate food insecurity will increase throughout the 2020s and severe food insecurity will plateau in most of sub-Saharan Africa, although some countries, such as Kenya, Ethiopia, and Ghana, will see improvements. Under our realistic scenario assumptions, food insecurity will fall in most other world regions, particularly South Asia and East Asia & the Pacific.



Number of people over the thresholds for moderate and severe food insecurity, by UN SDG regions over time, with a 3-year smoothing. Panel (A) shows the number of food insecure people with region totals stacked, to show global trends and totals over time. Panel (B) shows regional totals over time.

Examining the rate of food insecurity in a population, rather than the total number of food insecure people (See Fig. [2](#fig:rates)), we find that most progress has been made in East & Southeast Asia, while Central & Southern Asia is expected to see improvements in food insecurity throughout the 2020s. Other poor and middle income areas, including much of Africa, the Middle East and Latin America, have made only slight progress and are expected to continue reducing food insecurity 2030. In high-income countries in Europe and North America, the rates of people in at least moderate food insecurity do not tend to change, and rates of severe food insecurity are expected to increase slightly. Overall, the world has made steady but slow progress on reducing the rate of food insecurity.

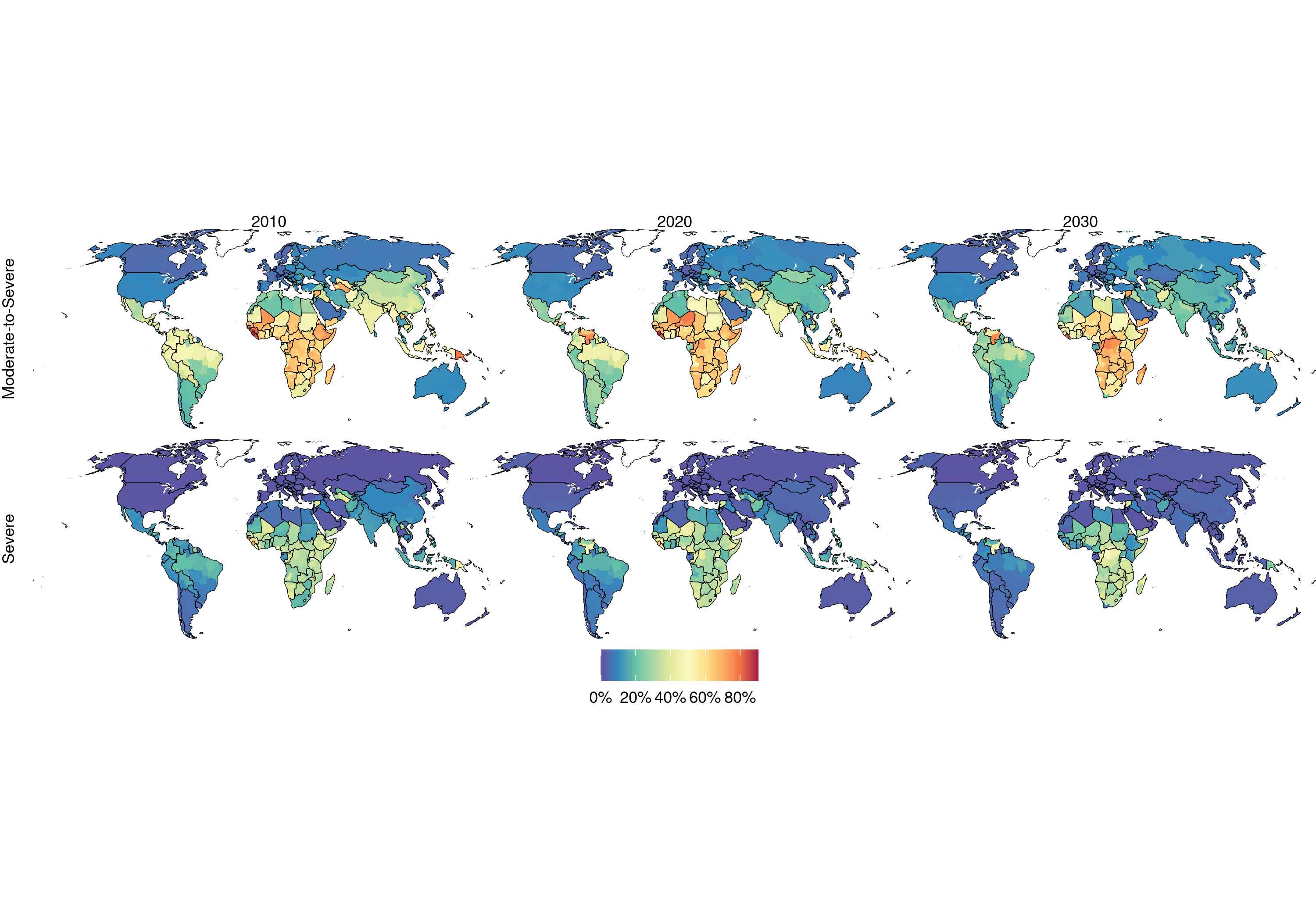


Percentage of the population over the threshold for moderate and severe food insecurity, by UN SDG regions over time, with a 3-year smoothing.

## Food Insecurity across Space

We find substantial heterogeneity in the global distribution of severe food insecurity across countries and subnational regions (see Fig. [4](#fig:map)). For the year 2020, mainland sub-Saharan Africa is the continent with with the highest rates of severe food insecurity, with more than 15% of people over the corresponding threshold in at last one subnational area in every country except Gabon and Equatorial Guinea. Outside of Africa, serious pockets of severe food insecurity also occur in Venezuela, Syria, Papua New Guinea, Yemen, and Afghanistan. In many large middle-income countries, severe food insecurity is also quite prevalent, with rates between 10% and 15% in northern Brazil, many central Asian and middle-eastern countries, as well as India and Indonesia.

The experience of at least moderate food insecurity is quite common in many parts of the world. In 2020, much of Africa, south and southeast Asia, and parts of Latin America had over 50% of the population living with moderate food insecurity. Even in highly developed economies such as Australia, the United States, and parts of eastern and southern Europe, over 10% of the population is above the threshold for moderate food insecurity.

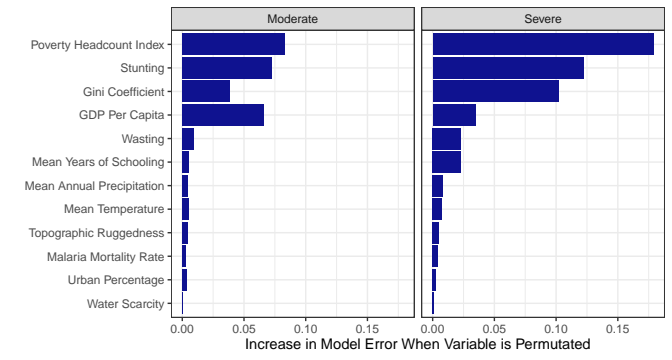


Spatial distribution of moderate and severe food insecurity for the years 2010, 2020, and 2030.

## Predictors of Food Insecurity

Overall, our model was able to predict food insecurity with high accuracy in the countries for which training data were available ( and for moderate and severe food insecurity, respectively). The relative importance of different variables as predictors of food insecurity can be assessed by examining the loss in forecast accuracy resulting from the exclusion of particular covariates.

For both moderate and severe food insecurity, the most important variable for predicting food security levels is the Poverty Headcount Index, followed by the rate of stunting. For predicting the rate of people with at least moderate food insecurity, GDP per capita appears particularly relevant, while for the rate of severe food insecurity, the country’s Gini coefficient appears more important.



Importance of each variable in predicting the percent of a population over thresholds for moderate and severe food insecurity. This is determined by the increase in the error when the variable is permuted before fitting the model.

The results of this analysis can be further explored in more detail on the [World Hunger Clock](https://worldhunger.io/), including statistics for each subnational administrative area.

# Discussion

This paper presents the first global estimates of food insecurity measures at a subnational level, with forecasts until the year 2030. We find that while food insecurity has been increasing and has likely worsened in 2020 due to the economic downturn associated with the coronavirus pandemic, our models predict that the number of people who experience food insecurity will decrease globally during the 2020s, largely because of progress in economic growth in south and east Asia. Nevertheless, many regions, most notably sub-Saharan Africa, are expected to see overall increases in the number of food insecure people.

Our analysis of the relative importance of different predictors for future food insecurity dynamics shows that economic conditions are the largest driver of global variation in rates of food insecurity. For both moderate and severe food insecurity, the poverty headcount index, the Gini coefficient, and GDP per capita appear as highly relevant predictors of the rate of food insecurity. The rate of stunting, an indicator of both sanitary conditions and chronic, long-term hunger is also an important predictor of food insecurity. However, other indicators of sanitary conditions, such as the rates of wasting and malaria mortality are less important as explanatory factors of changes in food insecurity. Geographic and climate factors, such as the percentage of urban population, topographic ruggedness and mean annual temperature or precipitation do not appear to be robust predictors of food insecurity at a global scale. Finally, while economic conditions are widely associated with food insecurity at a global scale, the fact that Europe and North America are predicted to experience little improvement in food insecurity outcomes until 2030, in spite of economic growth and continued reductions in poverty, suggests that there is an upper limit to the extent to which increasing economic development can improve food security outcomes.

These results contrast with the FAO’s predictions of increased rates of undernourishment, the other indicator for the monitoring of the SDG 2 target of ending hunger. This is in part due to our differing modeling approaches. While the SOFI report extrapolated recent trends in undernourishment and stunting, we modeled food insecurity based on interactions between projections in factors like GDP per capita, population, and the prevalence of stunting. Additionally, our diverging predictions for different SDG 2 indicators are not necessarily mutually exclusive: it is theoretically possible for undernourishment to increase while food insecurity, as measured by the FIES, decreases. The discrepancies between our projections and those of the FAO illustrate both the complexity of hunger as a social phenomenon, as well as the challenges in predicting future socioeconomic developments, where different modeling approaches may yield different forecasts.

When making predictions in countries where we do not have primary data, we assume that the relationship between variables like poverty, stunting or inequality, and hunger is the same across locations. We test the accuracy of these cases by using ten-fold cross validation at the country level, and we find reasonable accuracy (, MAE = 0.073 for moderate food insecurity and , MAE = 0.044 for severe food insecurity).

Additionally, when making predictions into the future, we assume that the long-term patterns of demographic change, urbanization, and development will maintain their trajectories for the next decade. At a global scale, this is almost certainly the case: it is highly unlikely that rates of fertility or economic growth shift suddenly globally. Nevertheless, at more local scales, sudden crises can lead to severe and unforeseeable increases in food insecurity, as crises in the previous decade in Yemen, Syria and Venezuela have shown.

While we find that food insecurity is expected to decrease overall by 2030, our model, which relies on middle-of-the-road assumptions, only expects the number of people experiencing moderate food insecurity to fall by 12% and the number of people experiencing severe food insecurity by 20%. This still leaves billions of people eating less than they should and nearly half a billion people going entire days without eating a decade from now, a number that falls far short of the SDG 2 goal of ending hunger. Moreover, decreases in food insecurity as measured by the FIES do not necessarily mean there will be improvements in other significant challenges such as poor dietary quality, micronutrient deficiencies, or obesity. Thus, while expected trends in correlates of food security give cause for optimism, there is still significant work to be done at the political level in pursuit of SDG 2.

# Methods

## Disaggregation

The data on the FIES collected by Gallup records many individual-level attributes from the respondents, including age, gender, wealth quintile, and whether they live in an urban or rural area. Using national data on these variables, a standard weighting scheme is created for each individual, based on the ratio of population probabilities to sample probabilities of an individual with those characteristics being selected . We use a similar methodology and calculate post-stratification weights at a subnational level. We use year-specific subnational estimates of population shares by gender, age, urbanization level, and, where available, wealth, and create a separate set of weights for each subnational area across all individuals. We use gridded data on population by age and gender from WorldPop. For urbanization, we use spatially explicit estimates published by Jones and O’Neill, and for wealth we use data on household wealth quintiles from Demographic and Health Surveys, where available. This methodology rests on the mild assumption that subnational differences in rates of food insecurity are driven by factors related to demographics, urbanization, and wealth rather than by other unobservable factors.

## Covariates

We model subnational rates of severe and moderate-to-severe food insecurity as a function of several covariates related to human development, food security, health, infrastructure, income levels and distribution, and the environment (See Table [[tab:covars]](#tab:covars)). We employ covariates that have subnational spatial resolution and for which projections are available up to the year 2030. For projections based on the Shared Socioeconomic Pathways (SSP) framework , we use the middle-of-the-road pathway, SSP2, and for climatological variables, we use forecasts based on the Representative Concentration Pathway (RCP) 6.0 .

Many of the variables require the harmonization of subnational historical data with projections at the national level. These include mean years of schooling, GDP per capita, as well as population. We use observed trajectories in the historical distribution of variables among subnational areas within a country to predict the future distribution of GDP, population, and schooling among subnational areas and disaggregate national-level future projections.

For health variables that have historically shown long-term trends and represent a rate of occurrence in a population, including stunting, wasting, and malaria mortality rate, we estimate the Annualized Rate of Change (AROC) for each subnational area to model these and to obtain predictions up to 2030. We calculate the rate of change between each pair of years in the dataset, and then use the mean rate of change over the period for which data are available, giving greater weight to more recent years, and applying that mean rate of change to estimate future levels. This method is used to obtain forecasts of the rates of stunting and wasting up to the year 2030.

For the climatic variables, temperature and precipitation, we combine historical observations with an ensemble of four bias-corrected simulations of the future climate from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP).

Finally, we adjust several variables, including GDP per capita, to account for the effects of the COVID-19 pandemic. For that, we use June 2020 GDP predictions from the World Bank, stunting and wasting estimates published in *The Lancet* in August 2020, as well as predictions for the poverty headcount index. For a detailed overview of the steps involved in processing and preparing each covariate, see the Appendix.

Covariates Included in the analysis

|  |  |  |
| --- | --- | --- |
| Name | Source | Scale |
| Urban Percentage |  | Subnational |
| Stunting |  | Subnational |
| Wasting |  | Subnational |
| Mean Years of Schooling |  | Subnational |
| GDP Per Capita |  | Subnational |
| Gini Coefficient |  | National |
| Poverty Headcount Index |  | National |
| Water Scarcity |  | Subnational |
| Mean Annual Precipitation |  | Subnational |
| Topographic Ruggedness |  | Subnational |
| Mean Temperature |  | Subnational |
| Malaria (*P. falciparum*) Mortality Rate |  | Subnational |

## Modeling

We fit two models for all subnational areas globally from 2010 to 2030: one for the percent of the population over the threshold of moderate food insecurity, and one for the percent of the population over the threshold of severe food insecurity. We use random forest regressions, which are known to perform well in the presence of non-linearities and interaction effects among predictors . A random forest regression involves generating a large number of decision trees, each under slightly different conditions and using different sub-samples of the data, and then aggregating the predictions of the decision trees. Given the bounded nature of our variables of interest, the random forest regressions are applied to the logistically transformed rates.

From a methodological point of view, key hyperparameters need to be elicited in the framework of random forest regressions. These include the number of decision trees to generate, the number of variables randomly selected as candidates for splitting a node, and the average number of observations in a leaf node. We calculate the Out-Of-Bag (OOB) error for different combinations of hyperparameters for the specifications used both for severe and moderate food insecurity. We choose the combinations of hyperparameters corresponding to the lowest OOB error. 5000 trees appear more than sufficient for the error rates to converge.

After fitting the models, we examine the importance of the individual covariates in explaining rates of moderate and severe food insecurity. We evaluate differences in Mean Average Error (MAE) from models including different explanatory variables corresponding to possible permutations of the covariates used.

## Model Validation

We obtain a MAE of 0.0029 for our model of severe food insecurity ( of 0.9978) and an MAE 0.0054 for our model of moderate food insecurity ( of 0.9981). To further validate the models, we carry out a ten-fold cross validation at the country level. We randomly divide the countries into ten groups, and fit the model recursively, using a group each time as out-of-sample validation sample. Then, we predicted the values for the ten percent that was left out using the model fit with 90% of the countries. By this validation metric, the modelling framework performed reasonably well, with an of 0.778 and an MAE of 0.073 for moderate food insecurity and an of 0.696 and a MAE of 0.044 for severe food insecurity. The Supplementary Materials provide a detailed description of the implementation and validation of the random forest models.

# Conclusion

# Supplementary Materials

# Overview of the FIES

We use raw microdata released by the FAO from 75 countries from different world regions and income categories (See Fig. [[fig:fies\_countries]](#fig:fies_countries).



Countries used in our model, shown in blue.

[fig:fies\_countries]

The microdata includes information about the probability that an individual is over the threshold for moderate and severe food insecurity, calculated by the FAO on the basis of the responses to the 8 FIES questions (See List [[itm:fies]](#itm:fies)).

The FIES is calculated using a Rasch model, which was developed by the psychometrics literature. The Rasch model assumes that each individual and their responses to the FIES questions can be placed on a one-dimensional scale of food insecurity, and that the log odds of a respondent answering affirmatively to one of the FIES questions is a linear function of the difference between the severity of the food insecurity experienced by theindividual and the severity of the item put forward in the corresponding question. The severity of each item and the respondent’s level of food insecurity can be estimated making use of maximum likelihood methods. The assumptions of the Rasch model hold up well for the data at hand, as evidenced by model fit diagnostics.

After separately estimating the Rasch model for each country, FAO develops a global reference scale for the severity of each of the eight questions by iteratively harmonizing the severity values in each country. These global reference points are calibrated against other surveys that ask questions similar to those in the FIES, such as the HFSSM and the ELCSA.

Based on the global reference scale, two thresholds are set based on questions 1 and 8: whether the respondent is worried they would not have enough food to eat because of a lack of resources, and whether there was a time when they went without eating for a whole day because of a lack of money or other resources.

1. During the last 12 MONTHS, was there a time when you were worried you would not have enough food to eat because of a lack of money or other resources?
2. Still thinking about the last 12 MONTHS, was there a time when you were unable to eat healthy and nutritious food because of a lack of money or other resources?
3. Was there a time when you ate only a few kinds of foods because of a lack of money or other resources?
4. Was there a time when you had to skip a meal because there was not enough money or other resources to get food?
5. Still thinking about the last 12 MONTHS, was there a time when you ate less than you thought you should because of a lack of money or other resources?
6. Was there a time when your household ran out of food because of a lack of money or other resources?
7. Was there a time when you were hungry but did not eat because there was not enough money or other resources for food?
8. During the last 12 MONTHS, was there a time when you went without eating for a whole day because of a lack of money or other resources? [itm:fies]

# Predicting Covariates Into the Future

## Annualized Rate of Change

For many of our covariates, to extrapolate recent trends to the future, we draw on historically annualized rates of change, which are used to project the future change in the variable. We use this method to estimate future rates of stunting, wasting, and malaria mortality, as well as to model future subnational shares of GDP and population. For each subnational area, , the rate of change (ROC) is calculated between each pair of adjacent years, , based on a value, , such that , available for each year,

.

Each ROC is weighted so as to give proportionally more weight to recent years. For a dataset with observations for 2010 to 2017, the corresponding weight for year *y* would thus be given by. The annualized rate of change (AROC) is calculated as. Finally, the projections, Proj, for each year up to 2030 are given by

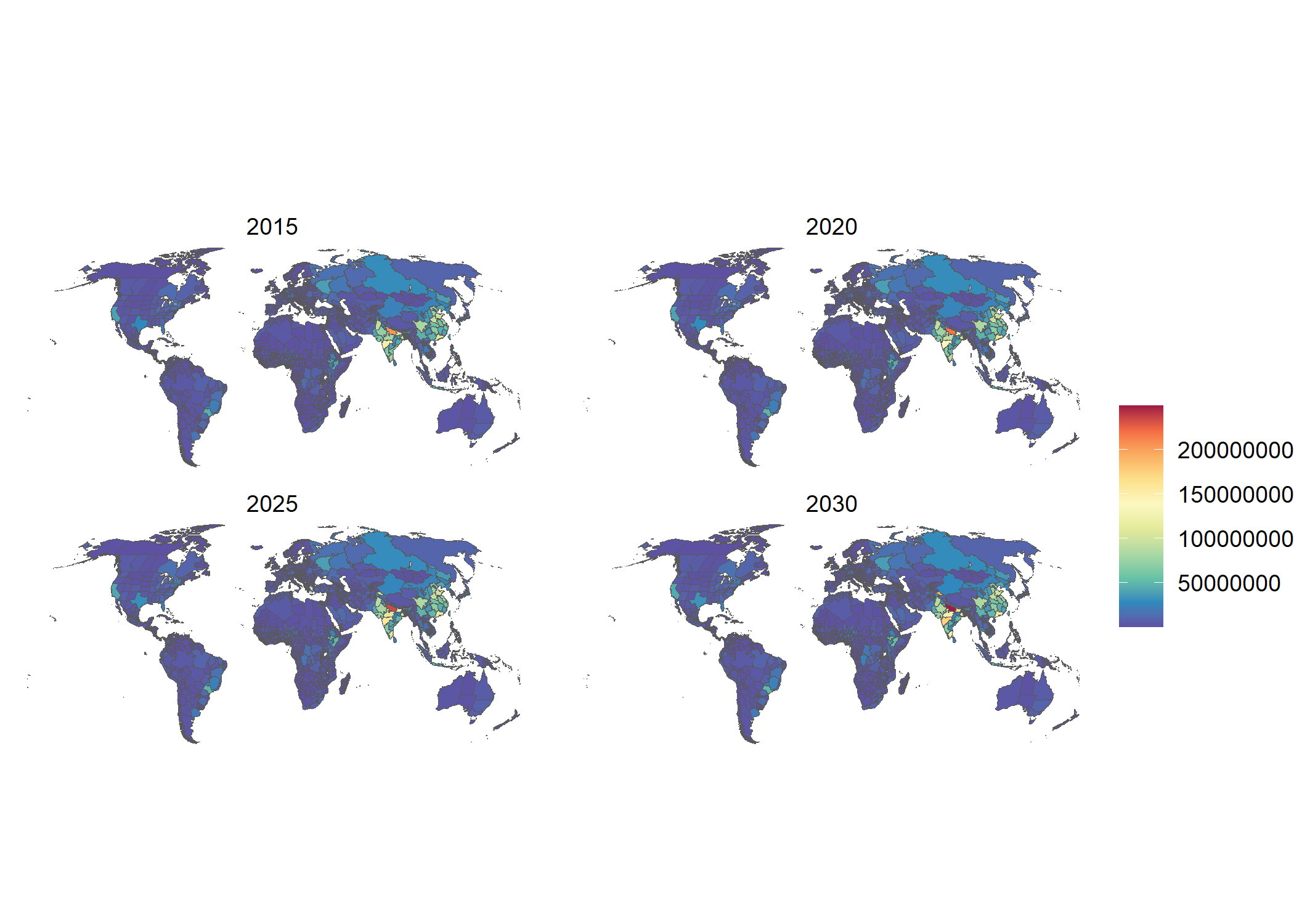
.

For cases where we model shares with a summing constraint, such as the regional shares of a variable measured at the national level, we rescale the shares to ensure they the constraint holds.

## Population

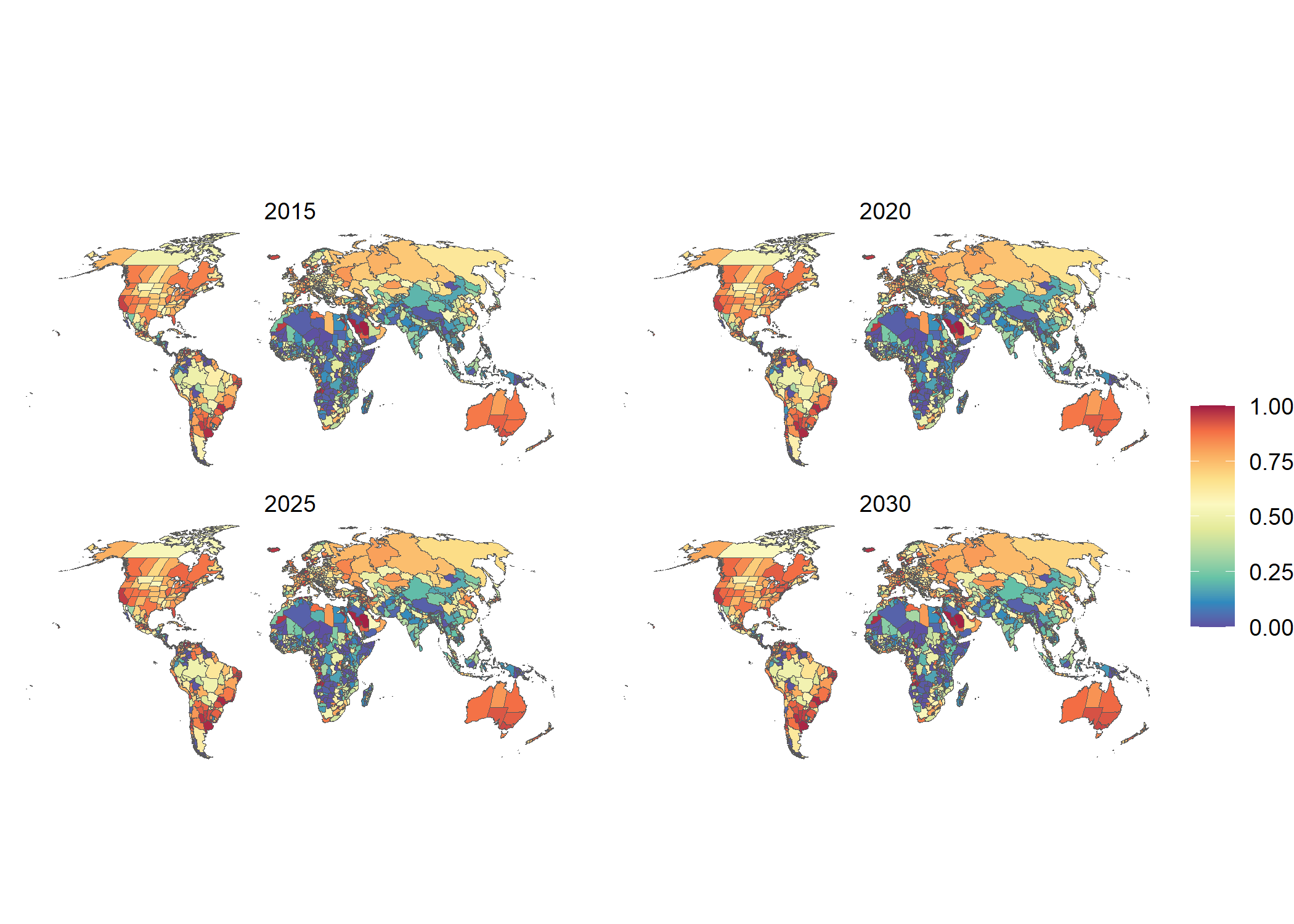
Although population is not a covariate in our random forest model, it is an important determinant of other covariates such as GDP per capita. Future estimates of population are also necessary to convert modeled future rates of food insecurity into total headcounts of food insecurity.

We project population by combining historical subnational data from the Subnational Development Database with national level population estimates for the 21st century . To account for within-country trends and changes in population distribution, we use the AROC method outlined in equations [[eqn:a]](#eqn:a) - [[eqn:d]](#eqn:d) to project each subnational areas share of national population totals. We then disaggregate the national totals given by KC et al. to estimate future subnational populations.

Population

## Urbanization rates

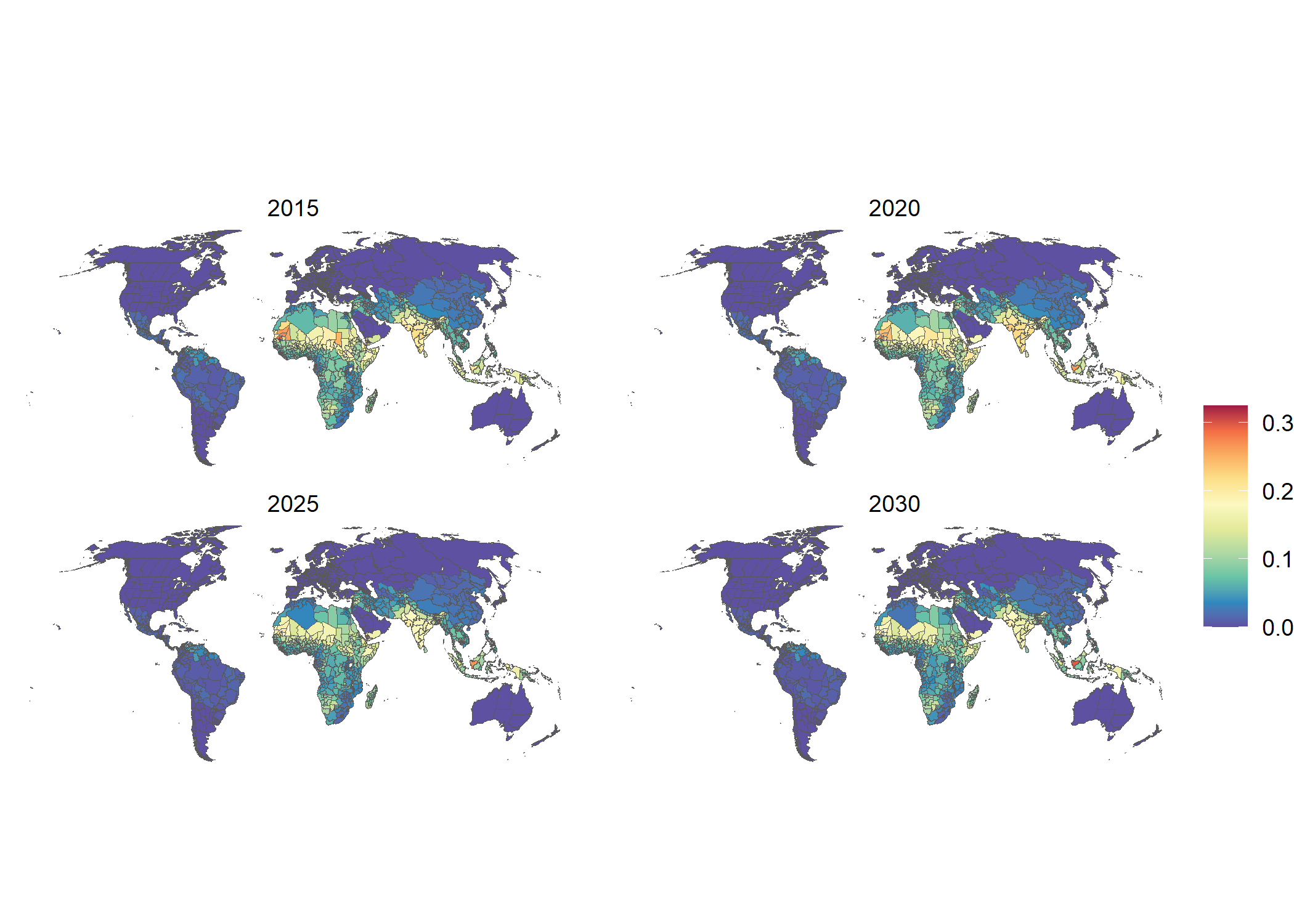
We draw our historical and future estimates of urbanization entirely from the spatially explicit scenarios of urban population by Jones and O’Neill, which provide a series of urbanization rate projections which are consistent with the Shared Socioeconomic Pathways. As with our other covariates, we use projections consistent with SSP2, the middle-of-the-road scenario.



Percentage of people living in urban areas.

## Wasting

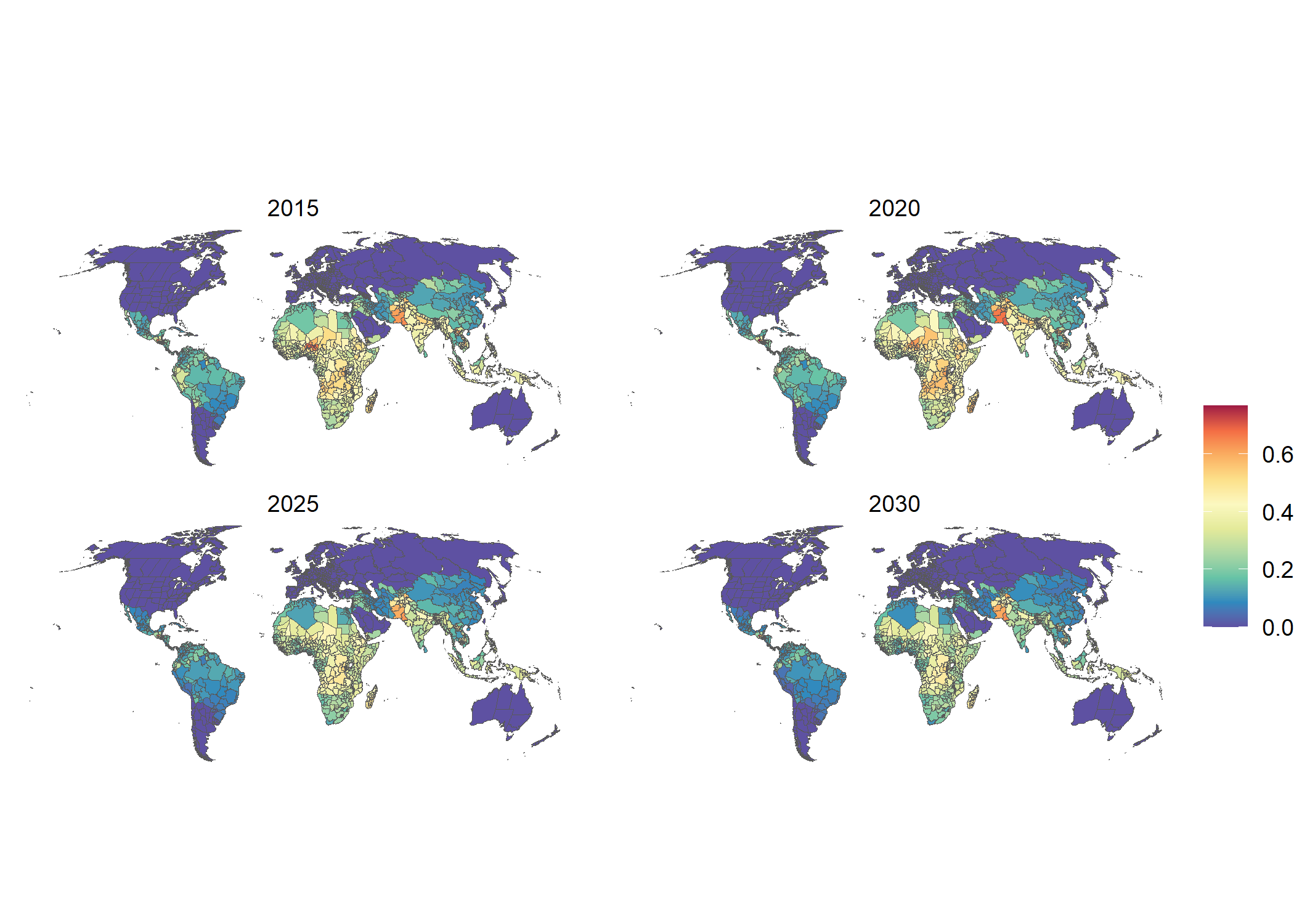
To estimate the prevalence of wasting for each administrative area in our dataset, we use data from the Local Burden of Disease project. For the years 2010 to 2017, we use the mean rate of wasting in each administrative area from the dataset. Higher-income countries that are not included in the dataset are assumed to have zero wasting. We project wasting for the years 2018-2030 using the AROC method, which the Local Burden of Disease group similarly use to estimate wasting for the year 2025. To account for the effects of the coronavirus pandemic on global rates of wasting, we assume that long-term trends in rates of wasting hold steady, after they increase globally by 14.3% in 2020, based on estimates by XXX. We model this impact as uniform across all countries where wasting occurs. We project the rate of wasting in each administrative area in 2021 as being the average of (a) the rate of change in 2020 accounting for the pandemic shock and (b) the predicted rate of change for 2022 obtained ignoring the effect of the pandemic.



Prevalence of wasting.

## Stunting

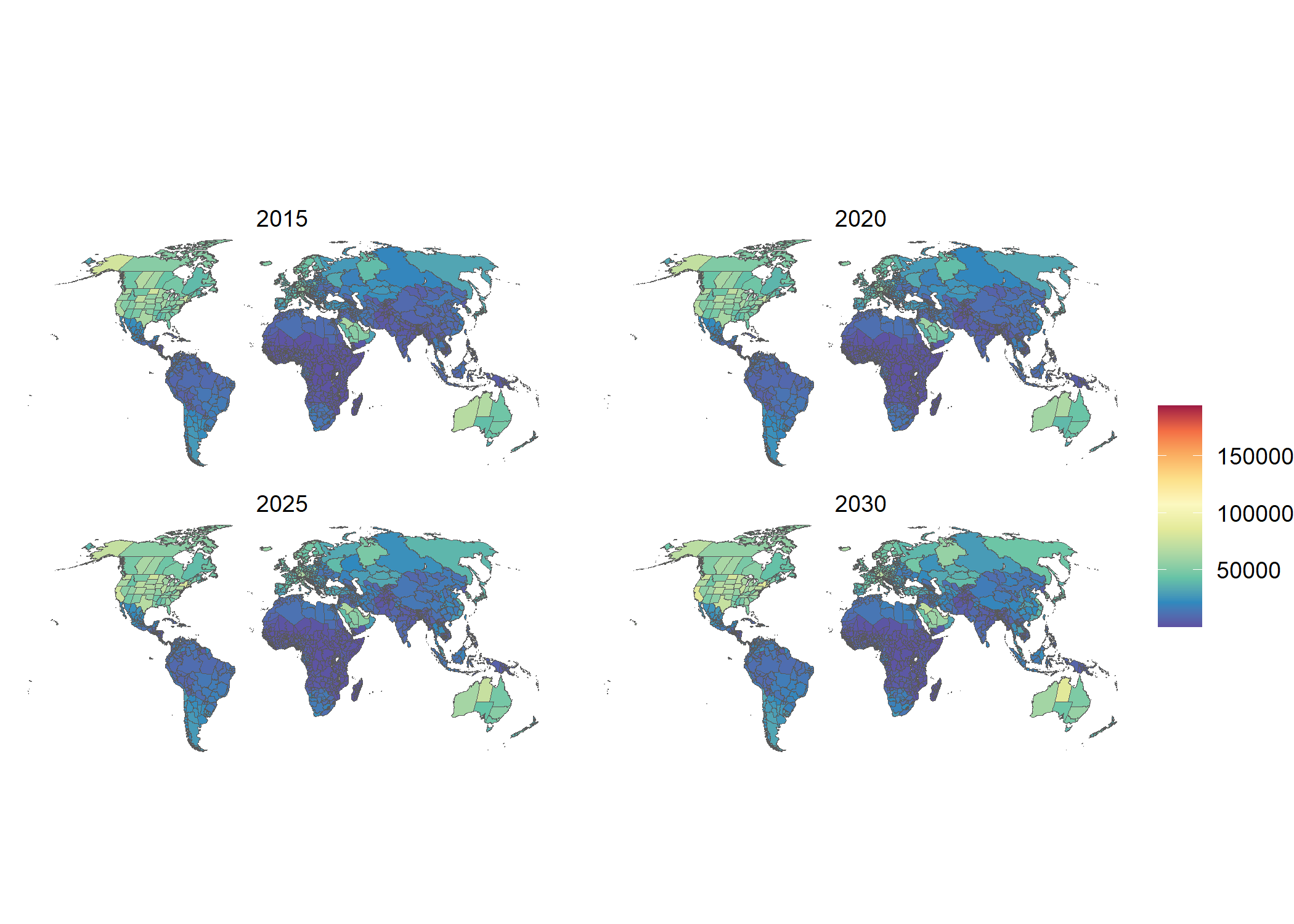
We model stunting, sourced from the Local Burden of Disease dataset, using the same methodology employed for wasting. We also assume a 14.3% increase in prevalence for the year 2020, as is done for wasting, although stunting is a long-term consequence of malnutrition and is not likely to be observable in a population with the same lags as wasting would be. We use stunting in our modelling framework as a proxy for chronic conditions of hunger, which are probably exacerbated by the coronavirus pandemic.



Prevalence of stunting.

## GDP Per Capita

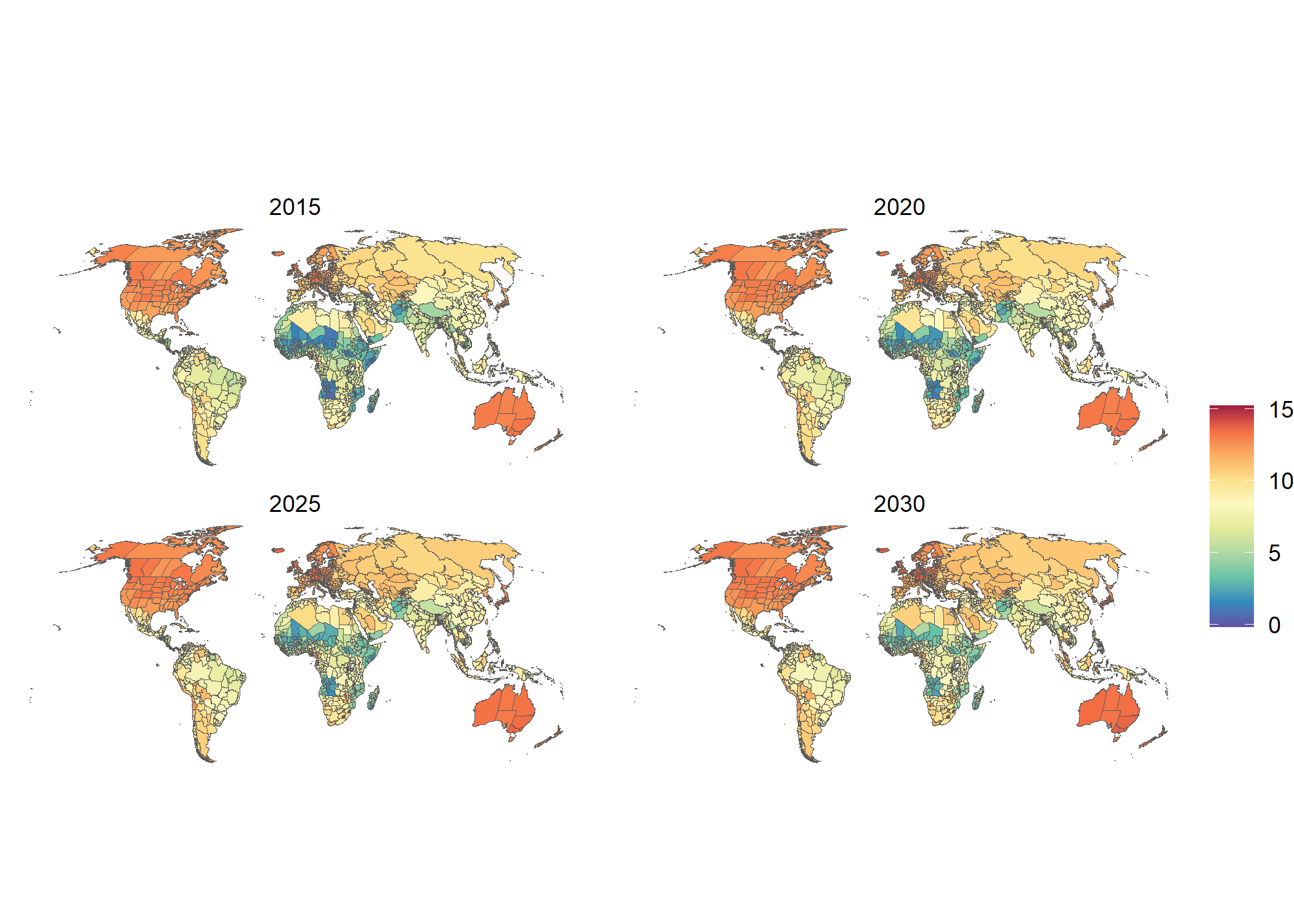
We measure global subnational GDP per capita using historic subnational estimates of Gross National Income (GNI) from the Subnational Human Development Database and forecasts of national GDP within the SSP framework from Dellink et al. . We source data on the mean country-level ratio of GNI to GDP from the World Bank and adjust the subnational estimates of GNI according to the ratio for each country in our sample. We calculate the share of national GDP for each subnational region, and use the AROC method to project each subnational area’s share of total country GDP up to 2030. National GDP totals are consistent with the projections of Dellink et al., but are disaggregated according to the historic patterns of subnational GDP shares sourced from the Subnational Human Development Database. We use nowcasts and forecasts of country-level changes in GDP growth for the years 2020 and 2021 to account for the impact of the coronavirus. After the shock created by the COVID pandemic, GDP is assumed to resuming its previous growth trajectory for the period 2022-2030.



GDP per capita.

## Mean Years of Schooling

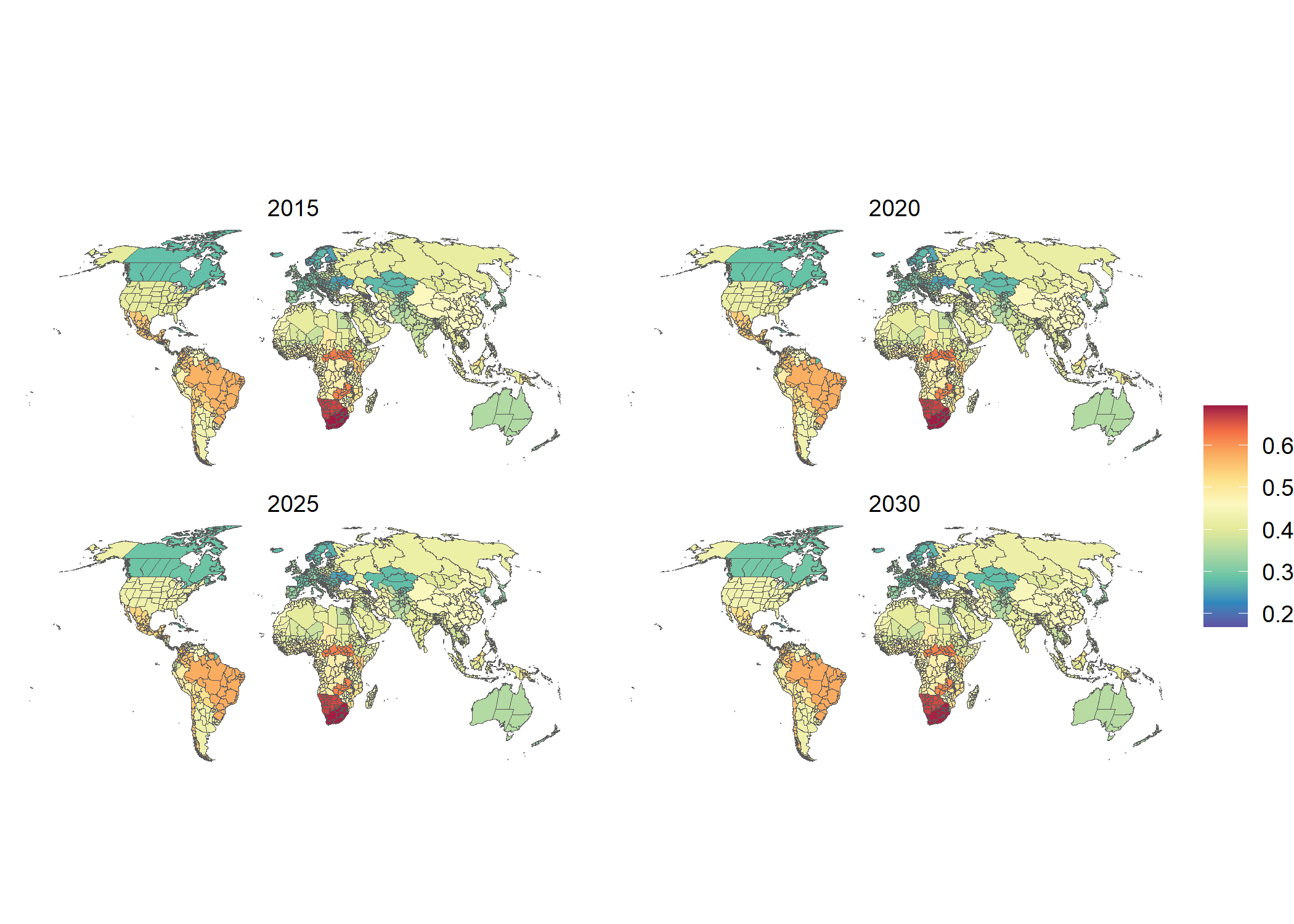
We use data on global subnational mean years of schooling from the Subnational Human Development Database and forecasts of national mean years of schooling from KC et al. . Using the subnational data, we calculate the difference in years of schooling from the national mean for each subnational administrative area, and use this value to disaggregate the future projections to a subnational level.



Mean years of schooling.

## Gini Coefficient

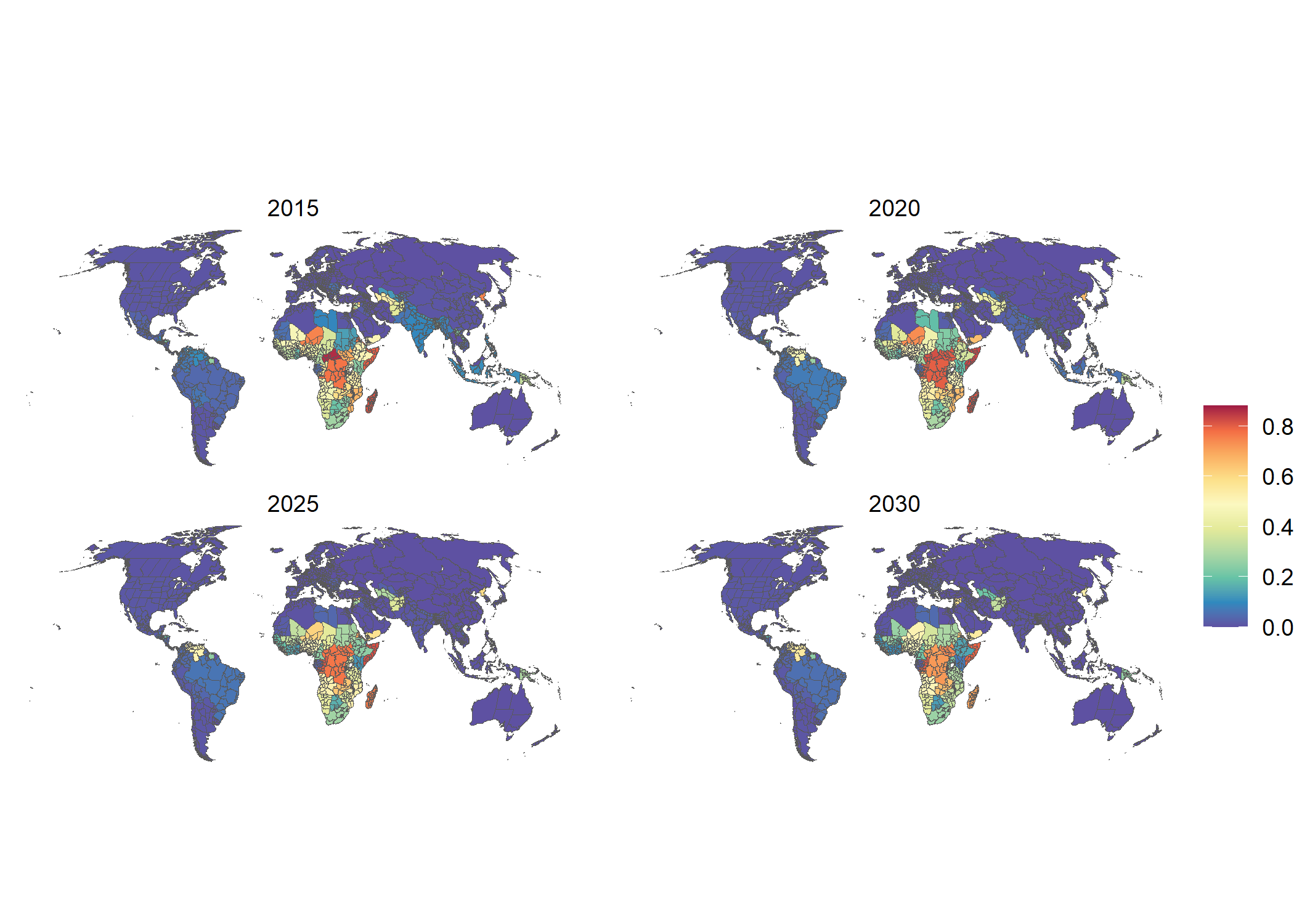
We use data on historical levels and projections of national income inequality from Rao et al. . These estimates are not disaggregated to a subnational level.



Gini Coefficient

## Poverty Headcount Index

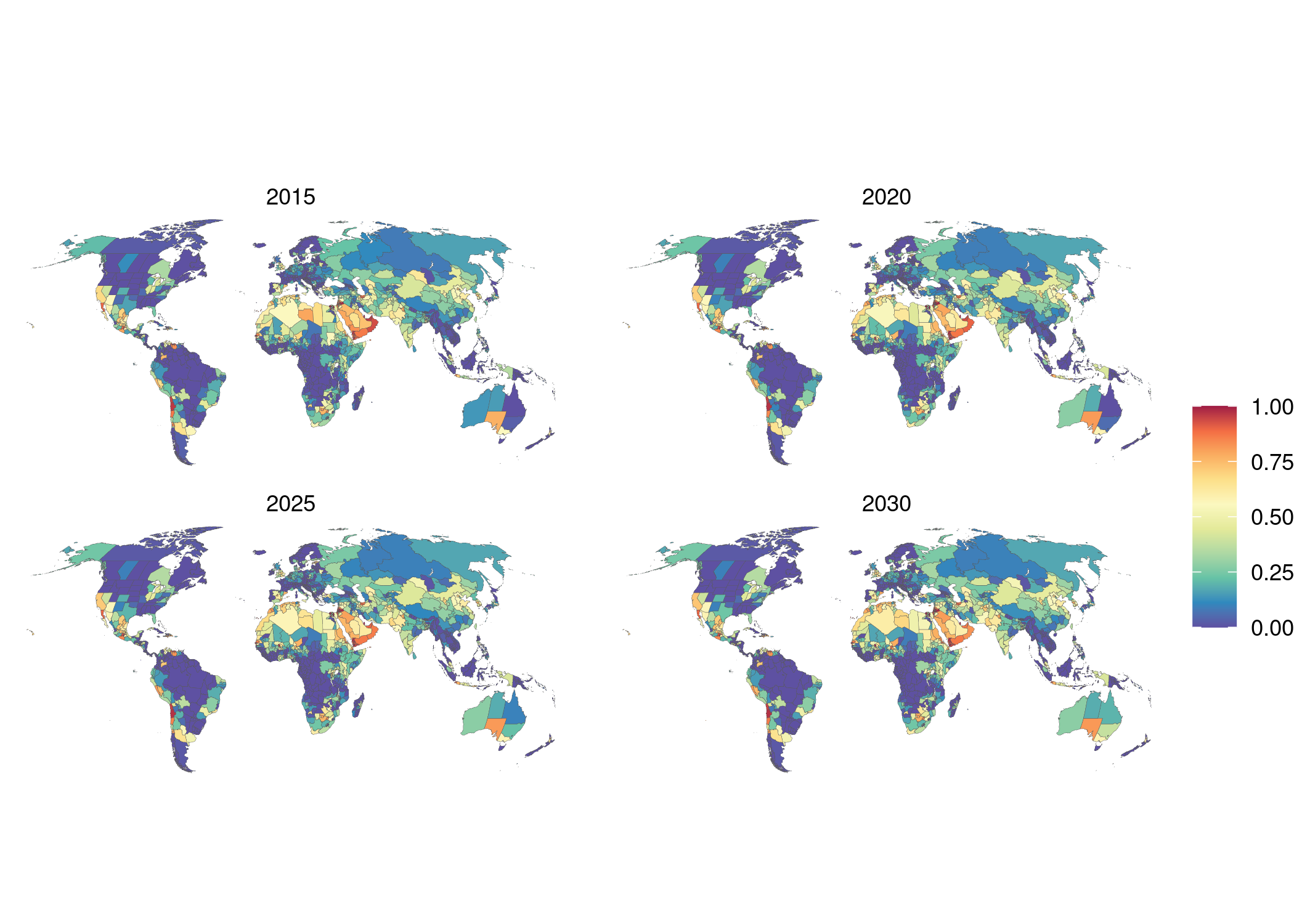
We used historical and projected data on the Poverty Headcount Index (using the standard threshold of $1.90 a day) from the World Poverty Clock by the World Data Lab, which in turn uses Povcal data sourced from the World Bank . These data are the version corresponding to the update in summer 2020 and include the expected impact of the COVID pandemic on global poverty dynamics. Using parameterized Lorenz curves in combination with mean income/consumption and population projections, poverty rates are projected.



Poverty Headcount Index

## Water Scarcity

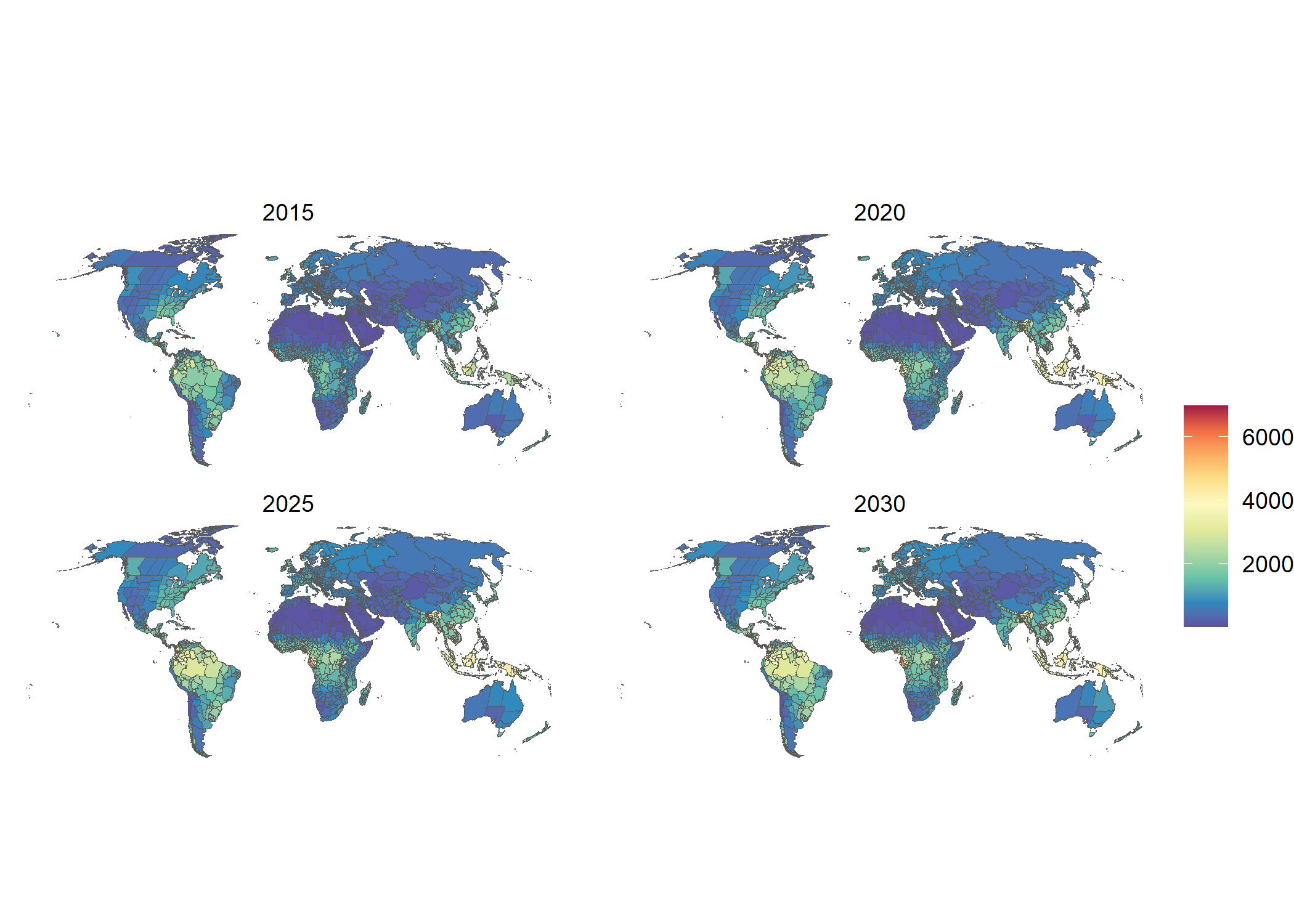
We use data for the water scarcity index (WSI) developed by . The index represents the ratio of the decadal averages in water demand to water supply and it assesses changes in water scarcity on a 0.5 by 0.5 degree global grid. Based on this index, we estimate the share of people living in areas with less than 1000m3 of water available per capita per year.



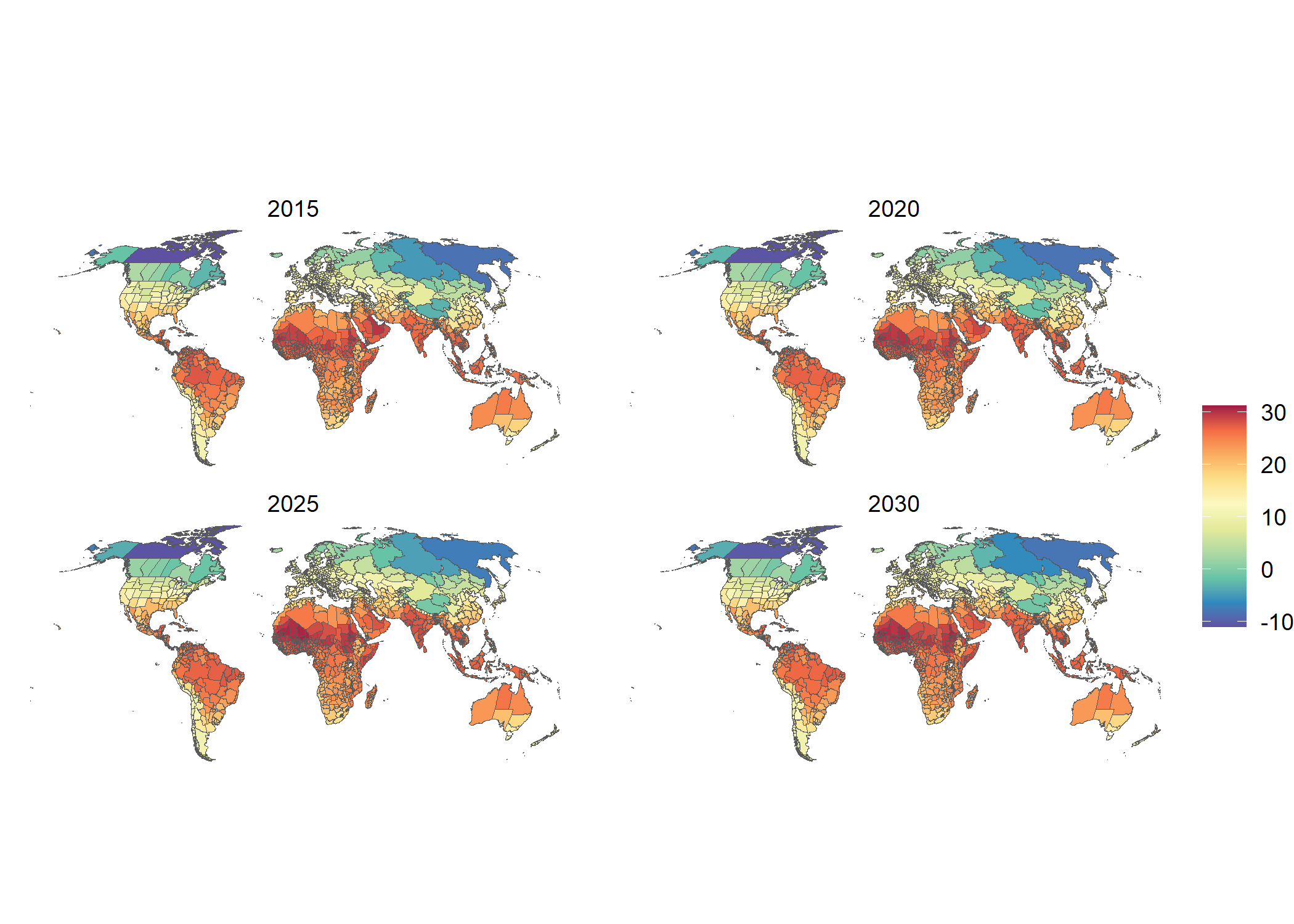
Water Scarcity Share

## Mean Annual Precipitation and Mean Temperature

For mean annual precipitation and temperature, we use data for the years 2010 to 2019 from the TerraClimate dataset , aggregated to each subnational area. For data for the years 2020-2030, we use projections from models from the Inter-Sectoral Model Intercomparison Project , in particular the mean of an ensemble of bias-corrected projections under representative concentration pathway (RCP) 6.0.



Mean annual precipitation



Mean temperature (Celsius)

## Topographic Ruggedness

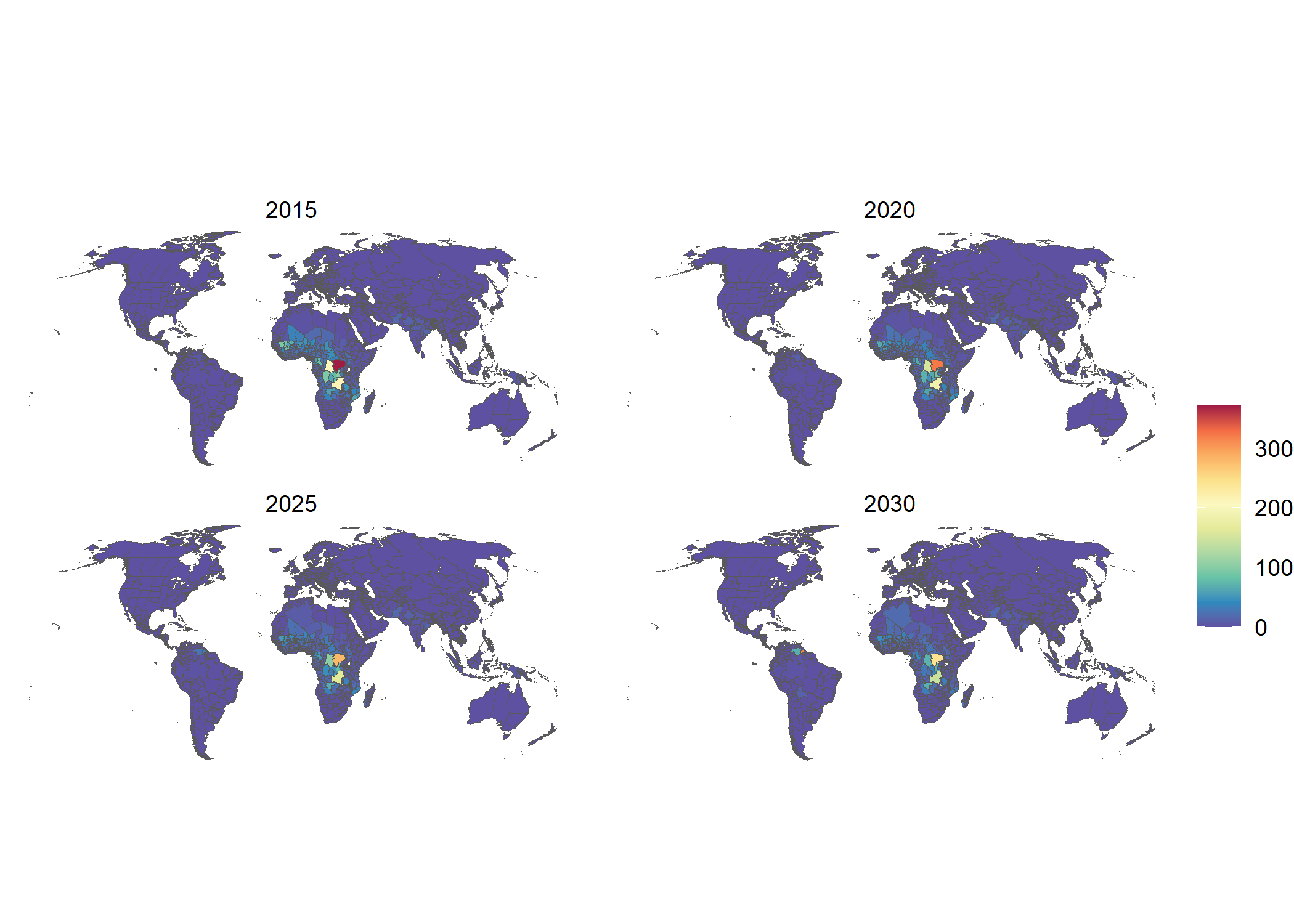
As a measure of accessibility, our models include the mean topographic ruggedness of each subnational area as a covariate. We obtain the variable using a gridded dataset of elevation from the USGS and calculate the index at each grid cell using the methodology from Riley et al. . We then aggregate the values for all the grid cells within each administrative area. The variable is constant in the estimation and projection period.



Topographic ruggedness

## Malaria (*P. falciparum*) Mortality Rate

We use data from *The Lancet* to estimate the mortality due to malaria (deaths per 100,000 population per year), and aggregate the information at the level of subnational administrative area. Projections of malaria mortality are based on the AROC method.

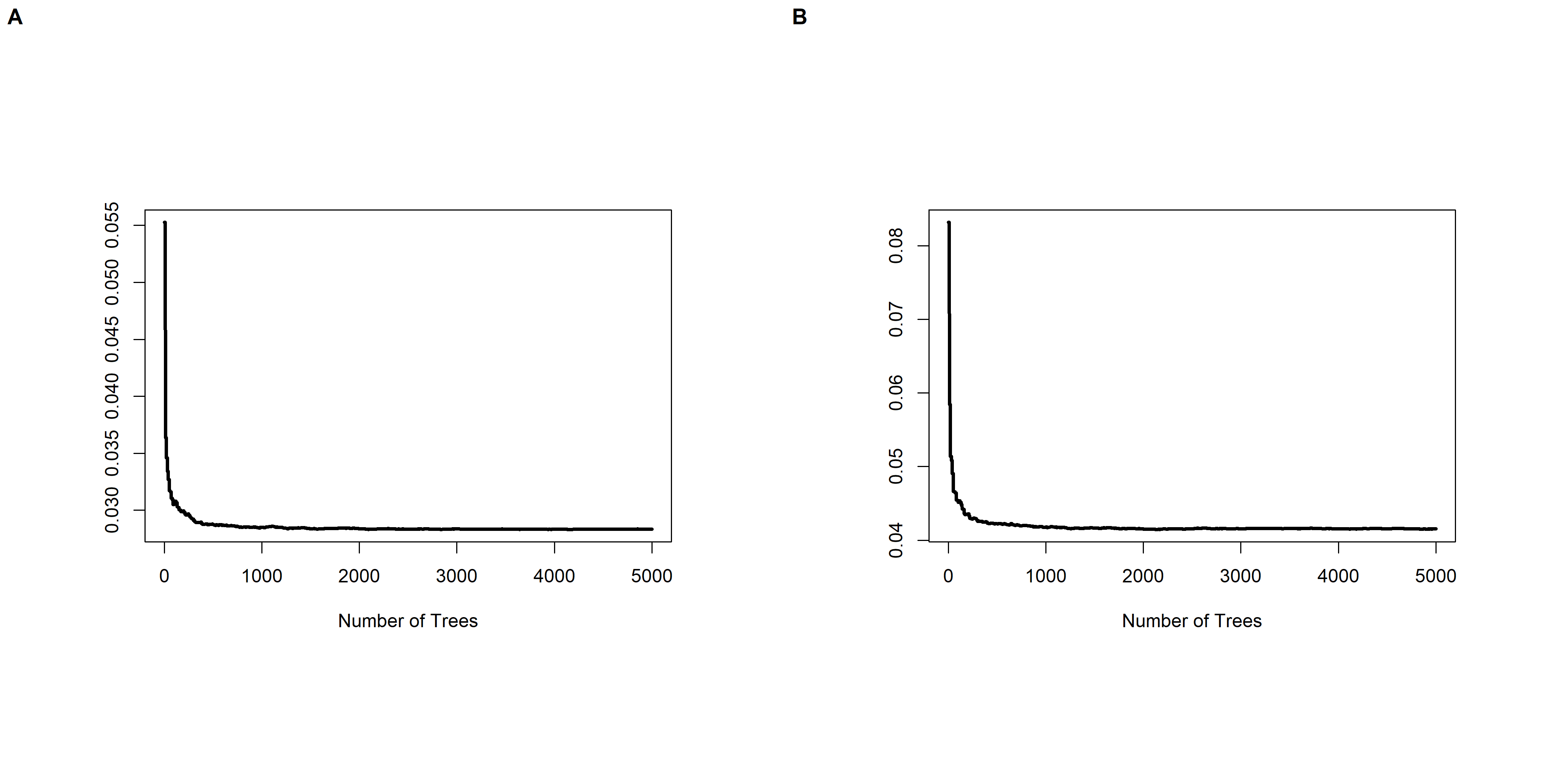


Rate of mortality due to P. falciparum Malaria

# Description, Implementation and Validation of the Random Forests Regression model

The RF regression is a machine learning method based on the creation of a large number of decision trees. By taking an ensemble of decision tree models, random forests introduce more variance and balance out the bias that is common to methods based on single decision trees . For each of the decision trees in a random forest model, observations are selected at random with replacement, a method known as bootstrap aggregating, or bagging, and features are also selected at random.

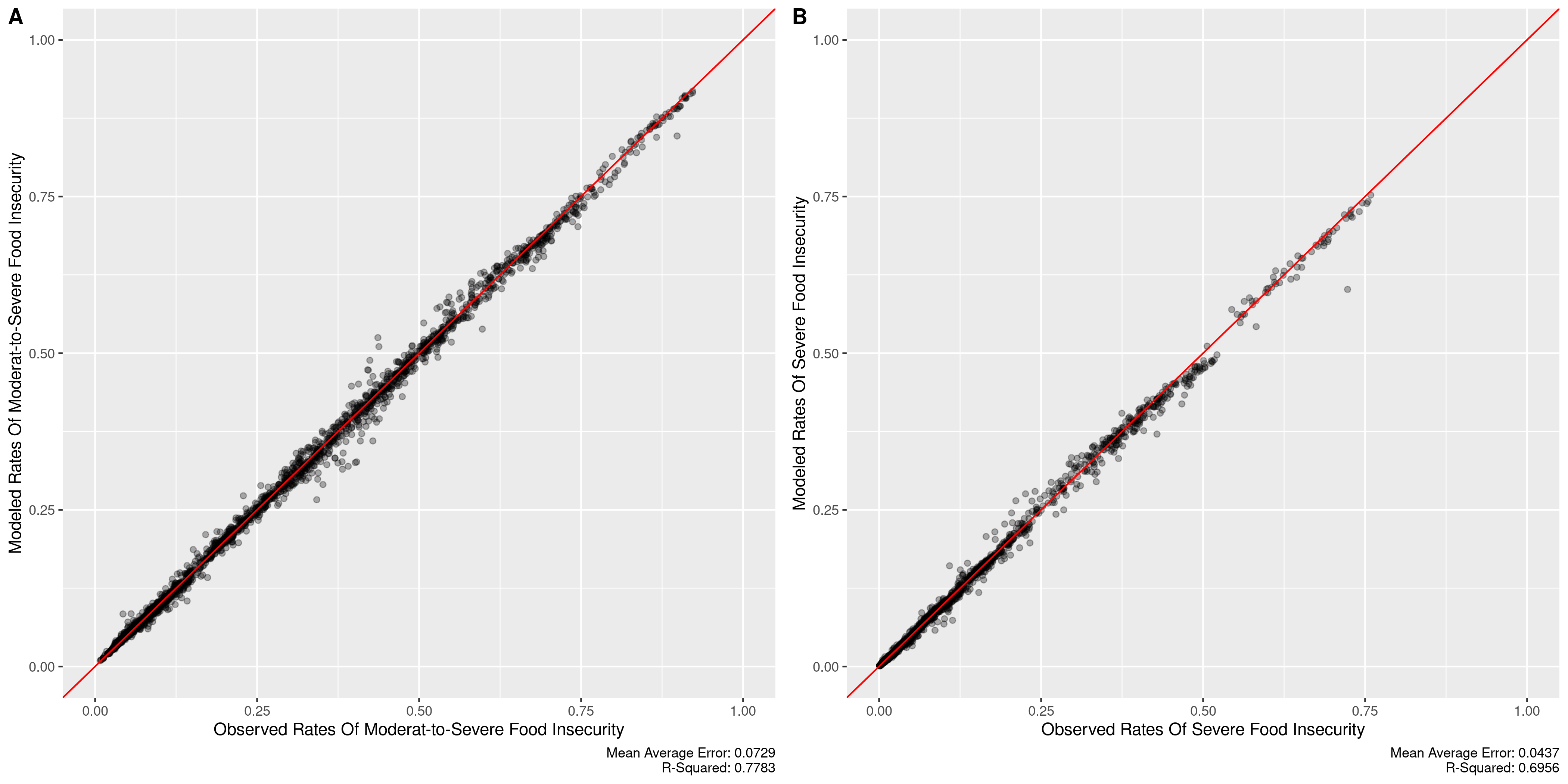
For the implementation of the method we use the R-package randomForestSRC by Ishwaran and Kogalur . Specifically, we apply the functions tune.rfsrc() to find the optimal combination of hyperparamters and the function rfsrc() to perform the RF regression models. After predicting rates of moderate-to-severe and severe food insecurity, we plot the modeled vs. observed rates and calculated the MSE and (See Fig. [[fig:rf\_in-sample]](#fig:rf_in-sample) and Fig. [[fig:rf\_out-sample]](#fig:rf_out-sample)) for both models. Additionally, we use the function vimp() to gain insights into the development of the error rates with increasing numbers of trees (See Fig. [[fig:rf\_error]](#fig:rf_error)) and the relative importance of individual variables (See Fig. [[fig:rf\_vimp]](#fig:rf_vimp)). To ensure that predicted rates are bounded by zero and unity, we apply a logit transformation of the dependent variable in the modelling step.



Error Rate of the RF regression model. Panel (A) shows the Moderate-to-Severe model and Panel (B) shows the Severe model.

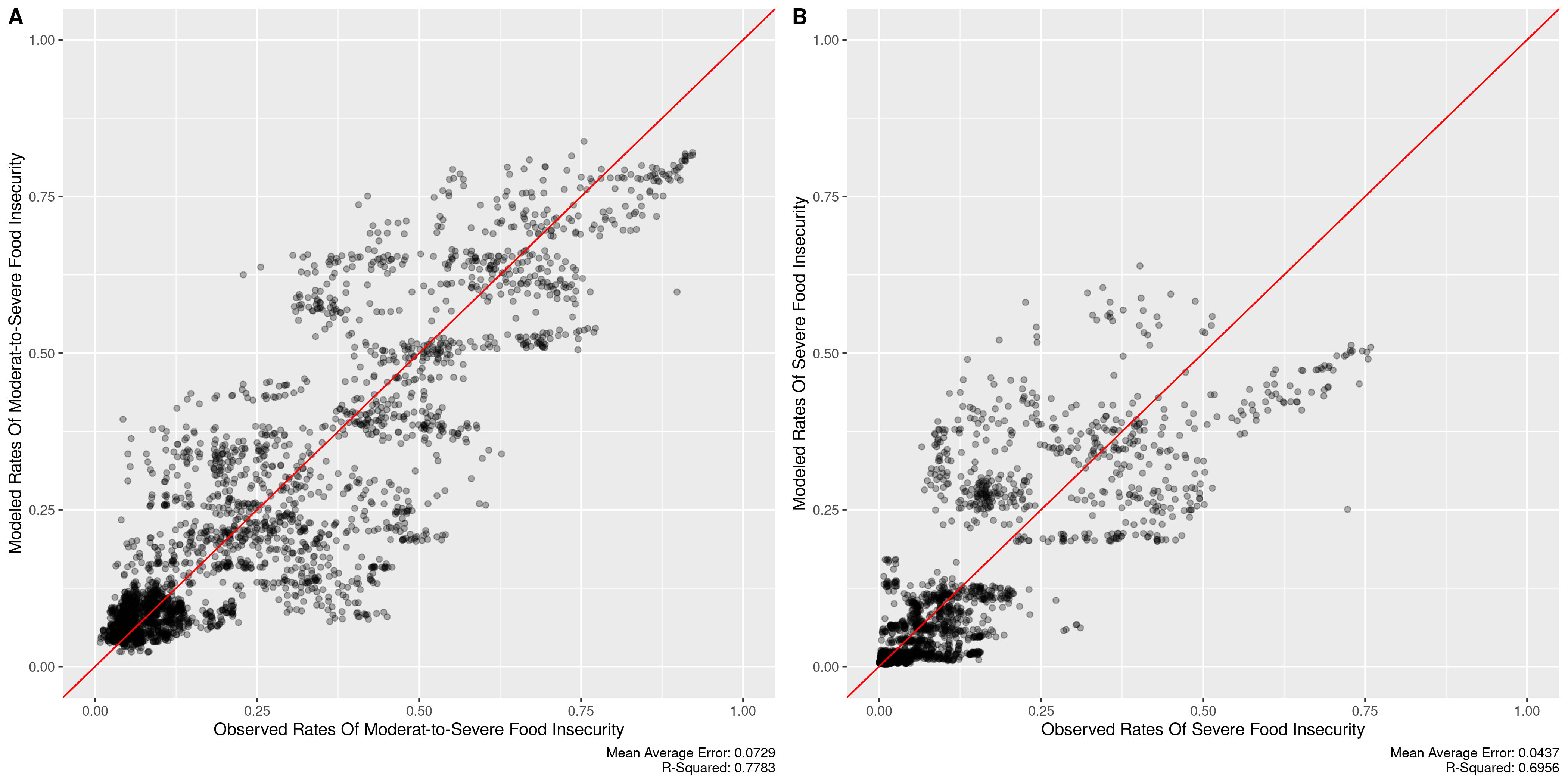
[fig:rf\_error]

In addition to the parameter that describes the number of trees to be created, two other important hyperparameters must be tuned. In the function tune.rfsrc() these parameters are called mtry and nodesize. The parameter mtry describes the number of variables randomly selected as candidates for splitting a node, and nodesize describes the average number of observations in a leaf node. The optimal pair of hyperparameters is found by choosing the combination with the smallest out-of-bag (OOB) error. The function tune.rfsrc() returns optimal values for mtry and nodesize. For the moderate-to-severe model the optimal combination is mtry = 12 and nodesize = 1 and for the severe model mtry = 10 and nodesize = 2 is found to be optimal.



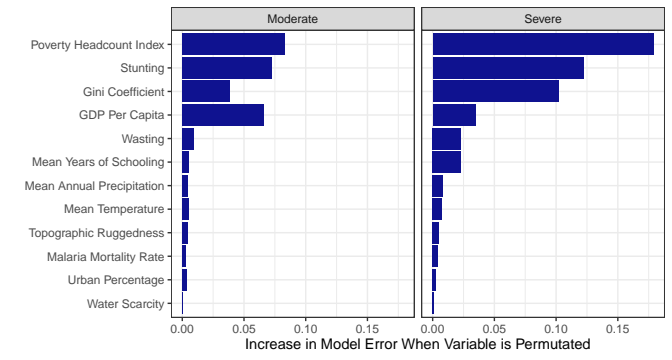
In-Sample Fit of the RF regression model. Panel (A) shows the Moderate-to-Severe model and Panel (B) shows the Severe model.

[fig:rf\_in-sample]



Out-Of-Sample Fit of the RF regression model with 75% training set and 25% test set. Panel (A) shows the model for moderate-to-severe deprivation and Panel (B) shows the severe deprivation model.

[fig:rf\_out-sample]



Variable Importance of the RF regression model. Panel (A) shows the Moderate-to-Severe model and Panel (B) shows the Severe model.

[fig:rf\_vimp]

The assessment of the importance of individual variables is based on the Breiman-Culter permutation variable importance information . This involves comparing the prediction error on the OOB data to the prediction error of OOB cases where a given variable is randomly permutated. The importance of an individual variable is the mean difference in prediction error between the perturbed and unperturbed error rate, across all trees.