CENTRAL EUROPEAN UNIVERSITY

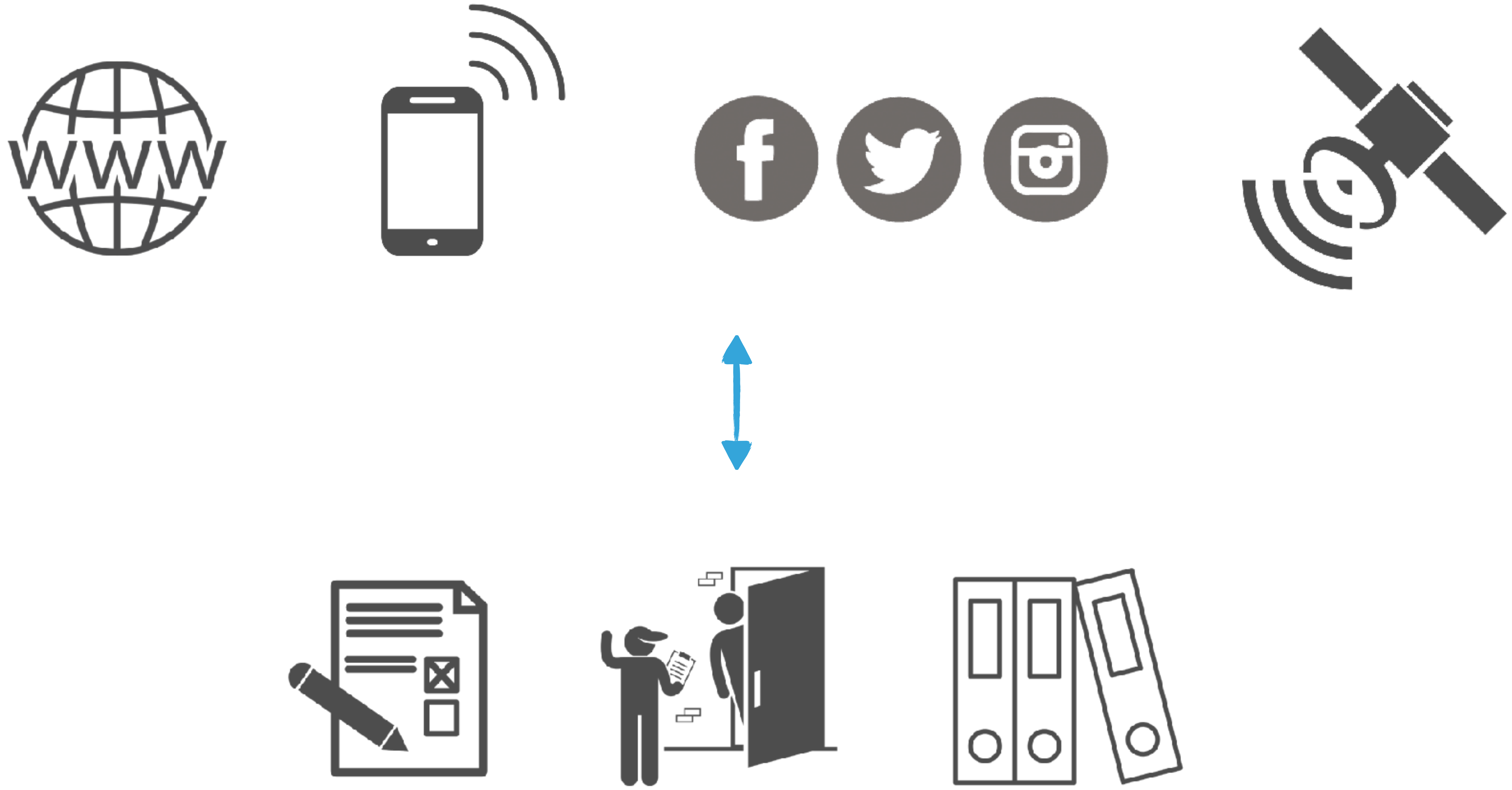
Computational methods for the Sustainable Development Goals
The case of food security

Elisa Omodei
Department of Network and Data Science
Central European University

Summer Institute in Computational Social Science | Amsterdam | 17 June 2024

ABOUT ME

- B.Sc. in Physics @ University of Padua, Italy
- M.Sc. in Applied Physics (Complex Systems) @ University of Bologna, Italy
- PhD in Applied Mathematics for the Social Sciences @ ENS Paris, France
- Postdoc @ Rovira i Virgili University in Tarragona, Spain
- Research Scientist @ UNICEF's Office of Innovation in New York, USA
- Lead Data Scientist @ UN World Food Programme's Research, Assessment & Monitoring Division in Rome, Italy
- Assistant Professor @ Department of Network and Data Science, Central European University in Vienna, Austria


WWW


1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

WHAT ARE THE SUSTAINABLE DEVELOPMENT GOALS?

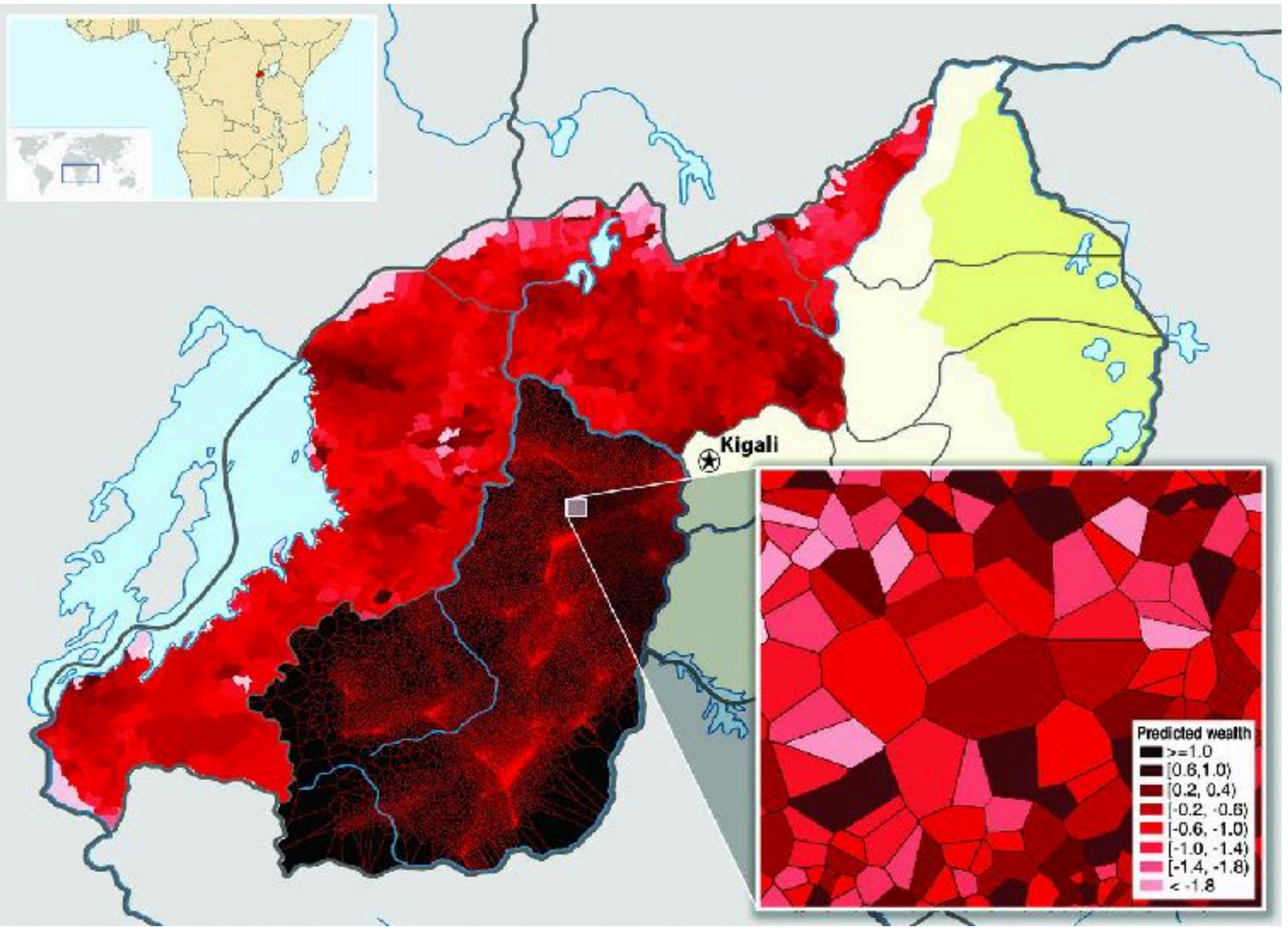
- 17 **goals** included in the "2030 Agenda for Sustainable Development" adopted by all UN Member States in 2015
- With the aim to "stimulate action over the next fifteen years in areas of critical importance for humanity and the planet"
- Each goal has a set of **targets**
- Each target has a set of **indicators** for monitoring progress

EXAMPLE: SDG 1

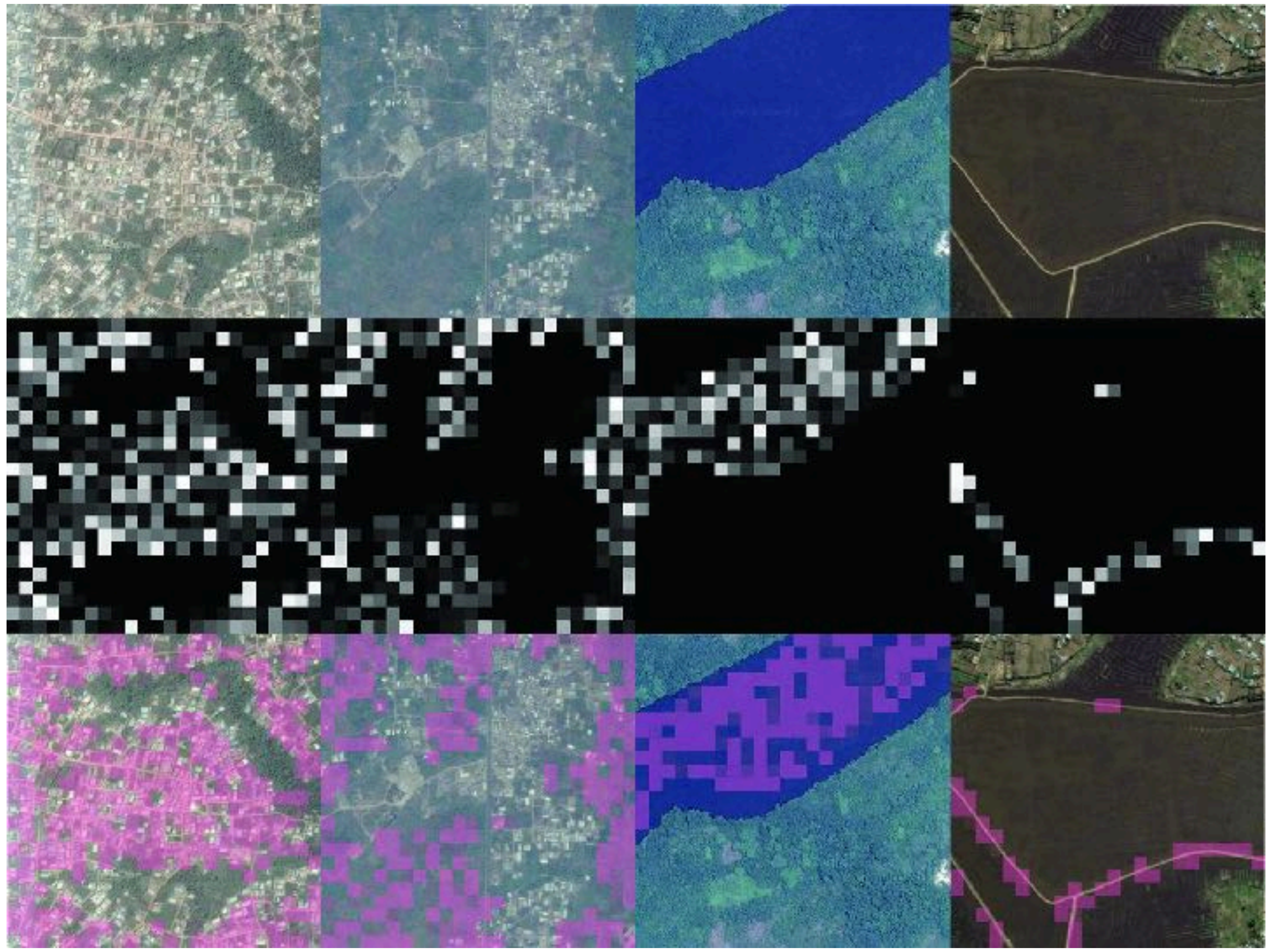
Goals and targets (from the 2030 Agenda)	*Indicators*
Goal 1. End poverty in all its forms everywhere	
1.1 By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day	1.1.1 Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)
1.2 By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	1.2.1 Proportion of population living below the national poverty line, by sex and age 1.2.2 Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
1.3 Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable	1.3.1 Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable

1 NO POVERTY

MAPPING POVERTY USING MOBILE PHONE & SATELLITE DATA



Blumenstock *et al*, Science 350.6264 (2015)



Jean *et al*, Science 353.6301 (2016)

1 NO POVERTY

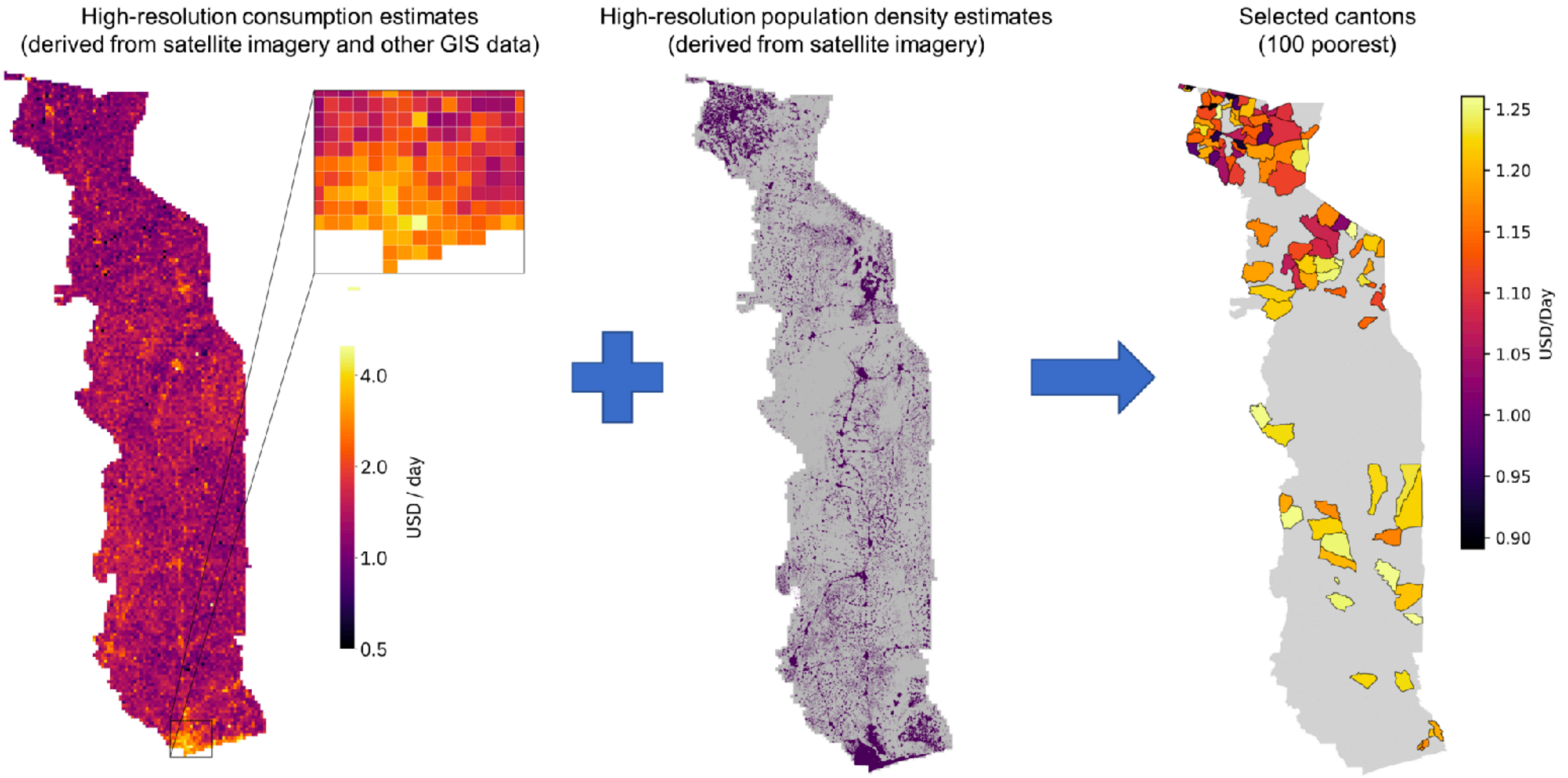
PRIORITIZING THE MOST VULNERABLE

Togo's Covid-19 innovative social assistance program "Novissi"

Aiken *et al*, Nature 603, 864–870 (2022)

PRIORITIZING THE MOST VULNERABLE

Stage 1: Prioritizing the poorest villages and neighborhoods

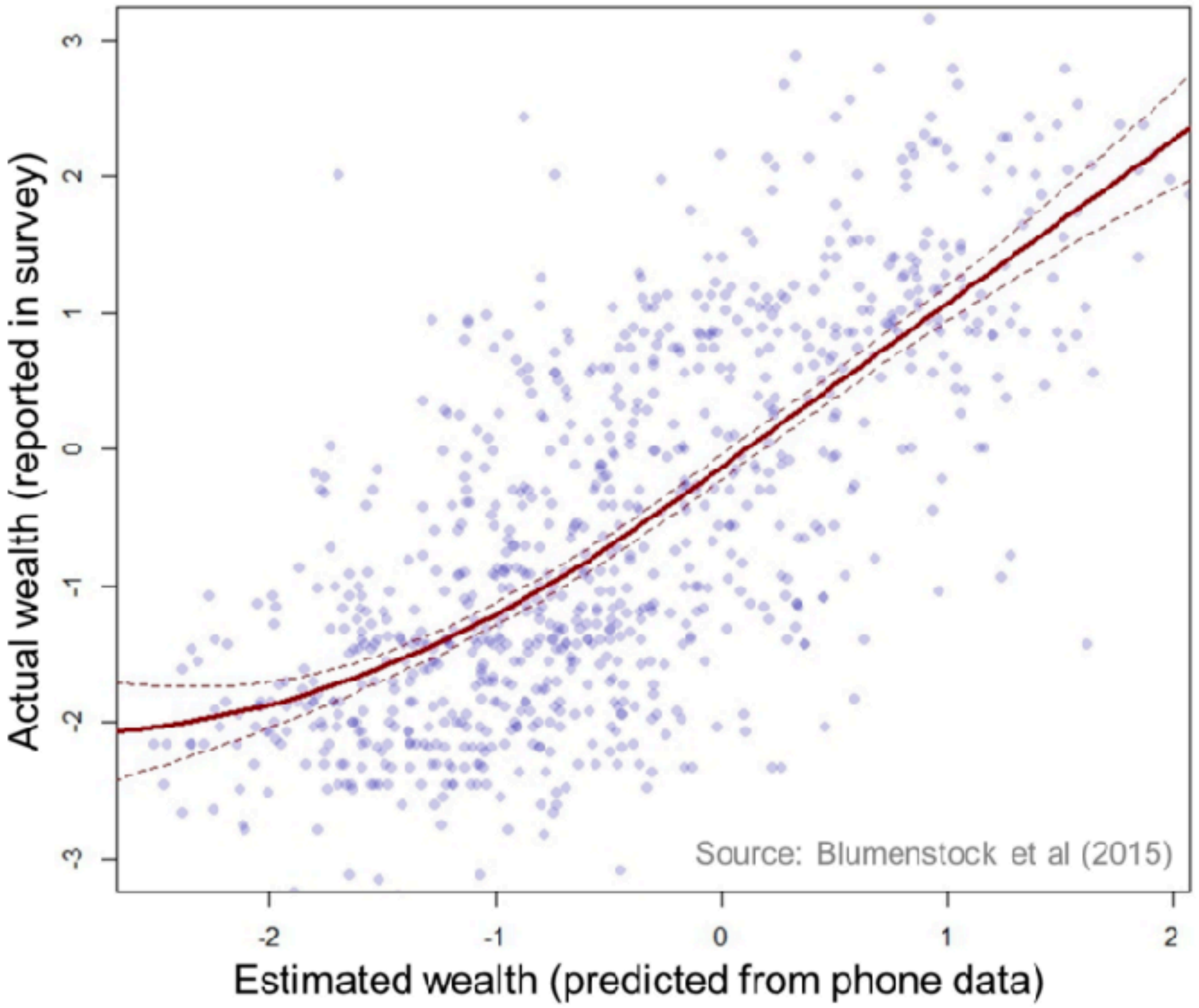


Aiken *et al*, Nature 603, 864–870 (2022)

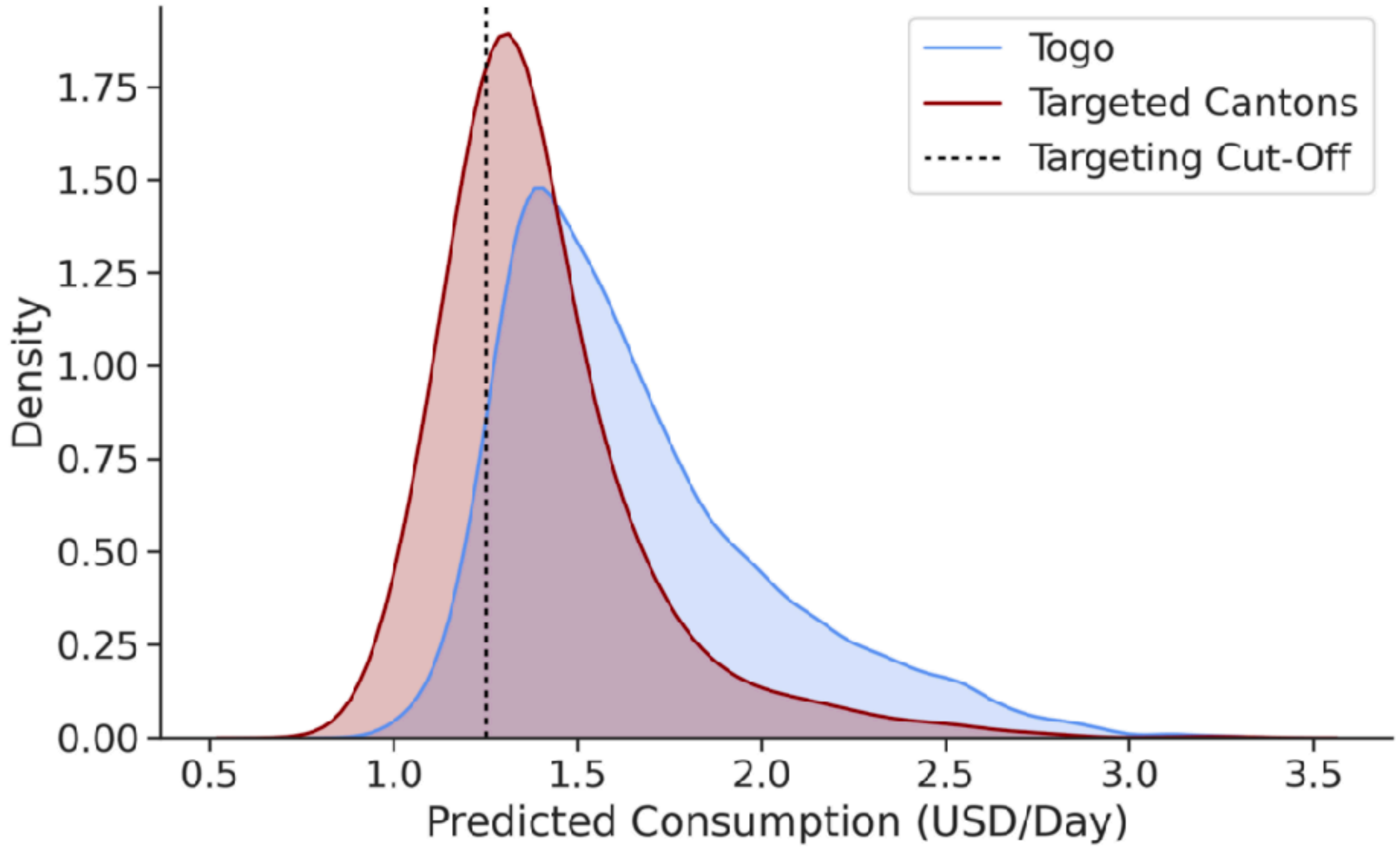
PRIORITIZING THE MOST VULNERABLE

Stage 2: Prioritizing the poorest individuals in the poorest villages

Predicting wealth from mobile phone metadata

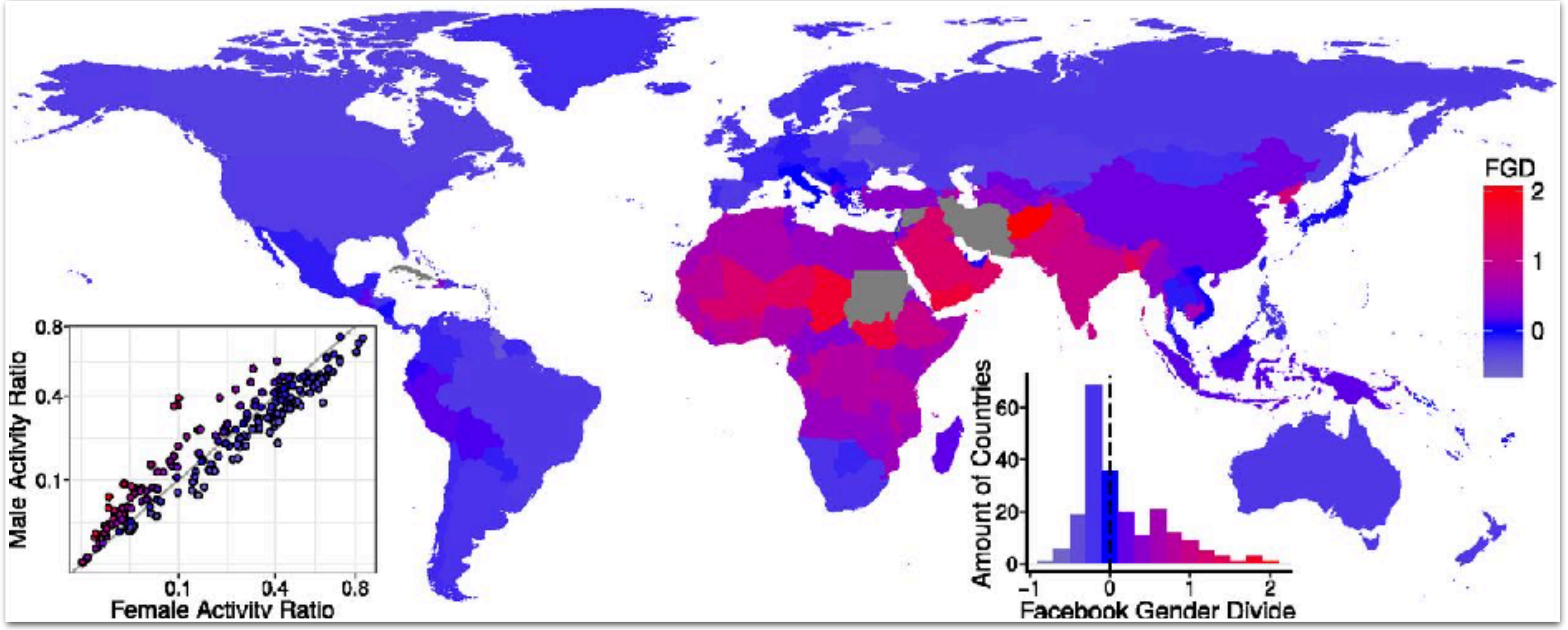


Targeting the poorest mobile subscribers

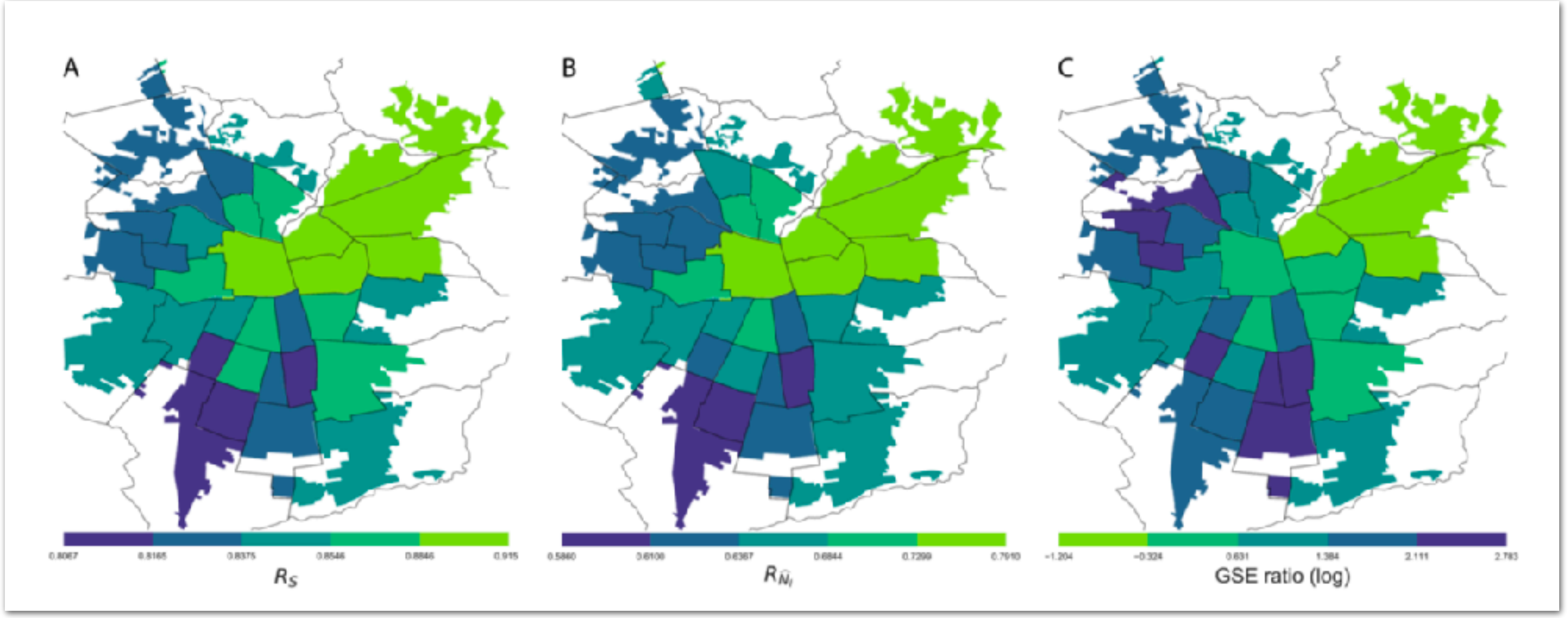


Aiken *et al*, Nature 603, 864–870 (2022)

IDENTIFYING GENDER INEQUALITIES

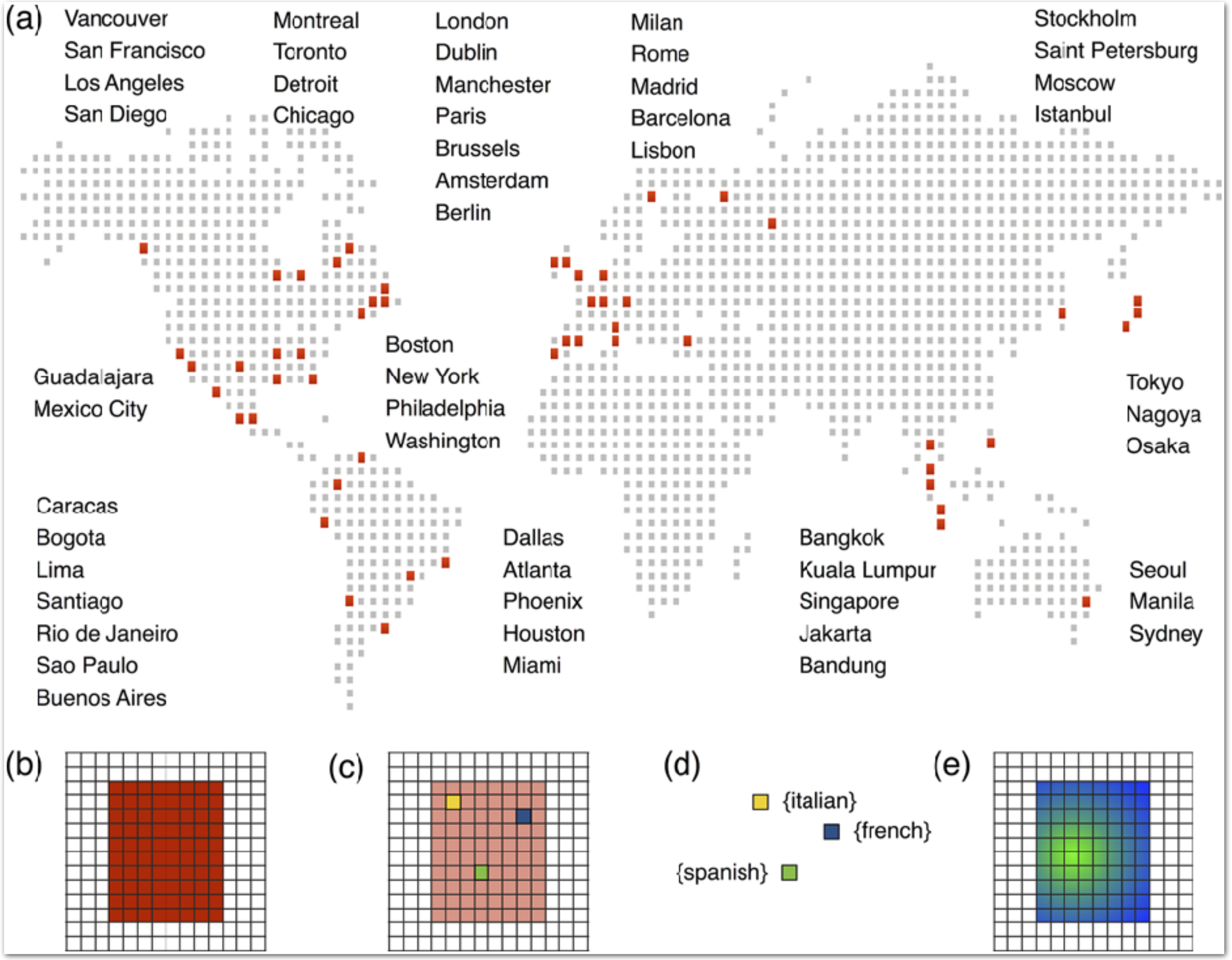


Garcia *et al*, PNAS 115.27 (2018)

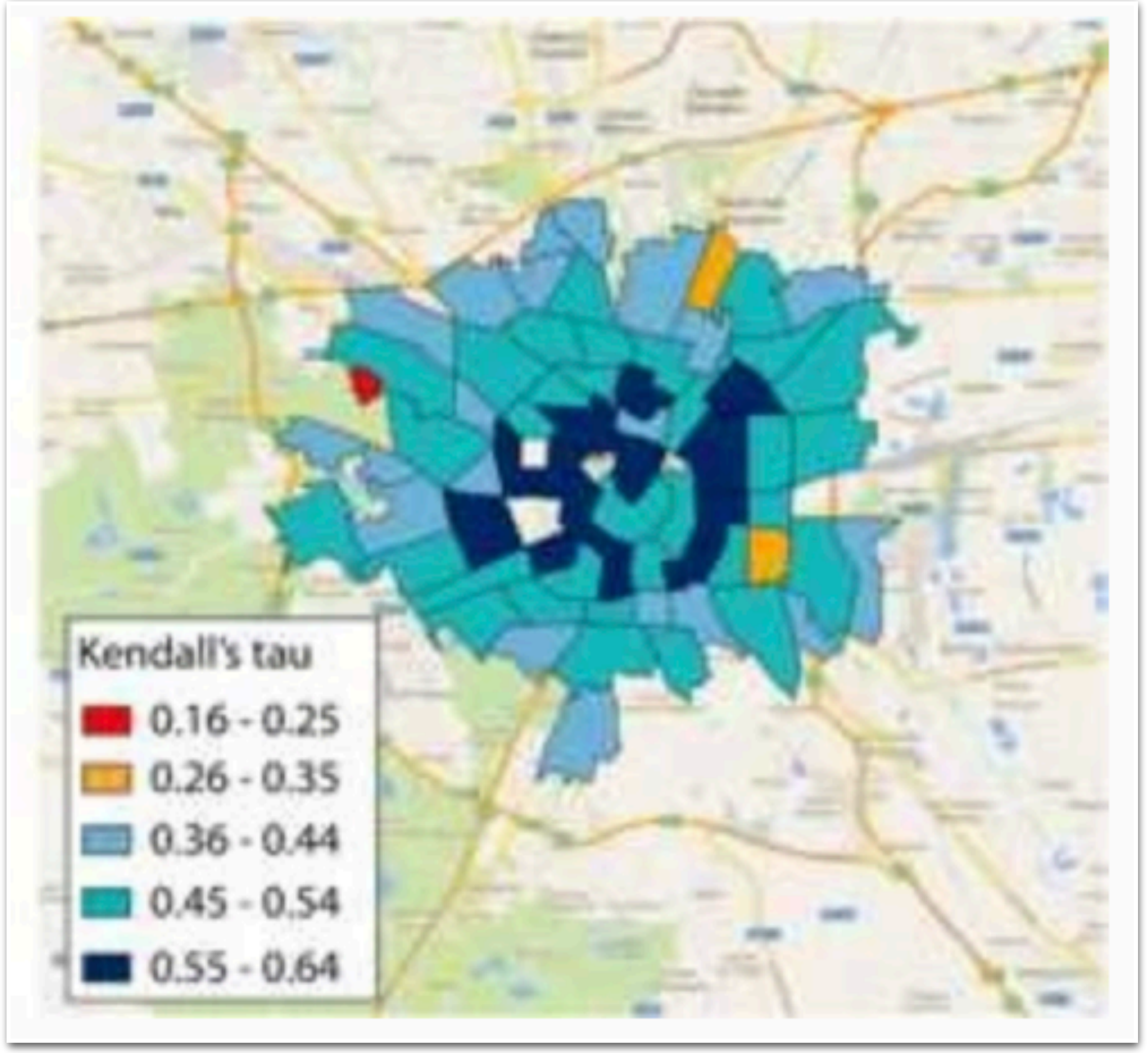


Gauvin *et al*, Humanities and Social Sciences Communications 7.1 (2020)

QUANTIFYING IMMIGRANT INTEGRATION IN CITIES



Lamanna *et al*, PloS ONE 13.3 (2018)



Bajardi *et al*, EPJ Data Sci. 4, 3 (2015)


1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
SUSTAINABLE DEVELOPMENT GOALS

WHAT IS FOOD INSECURITY?

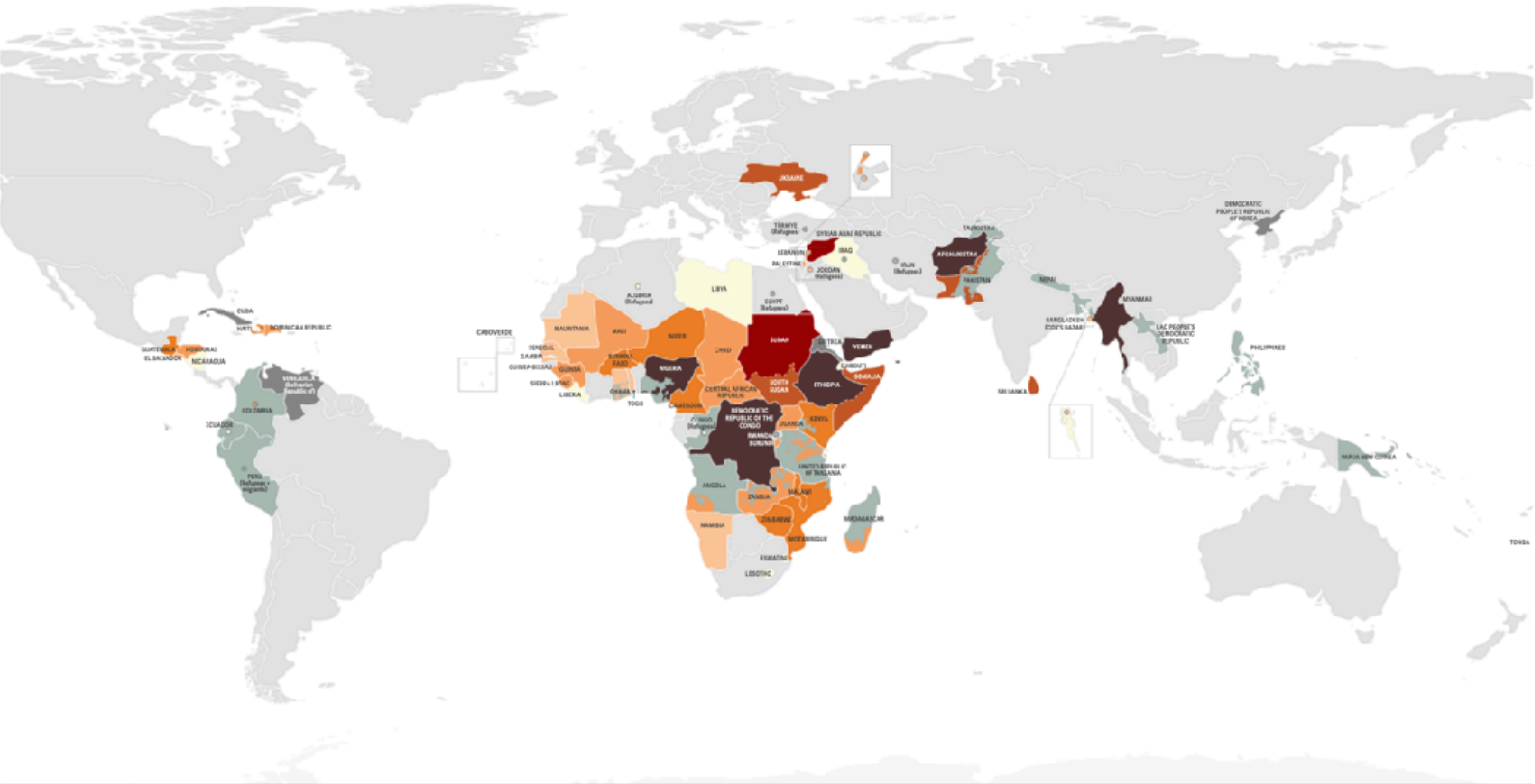
"FOOD SECURITY EXISTS WHEN ALL PEOPLE, AT ALL TIMES, HAVE PHYSICAL AND ECONOMIC ACCESS TO SUFFICIENT, SAFE AND NUTRITIOUS FOOD"

World Food Summit (1996)

258M ACUTELY FOOD INSECURE PEOPLE IN 2022

MAP 1.4

Numbers of people in Crisis or worse (IPC/CH Phase 3 or above) or equivalent in 58 countries/territories in 2022

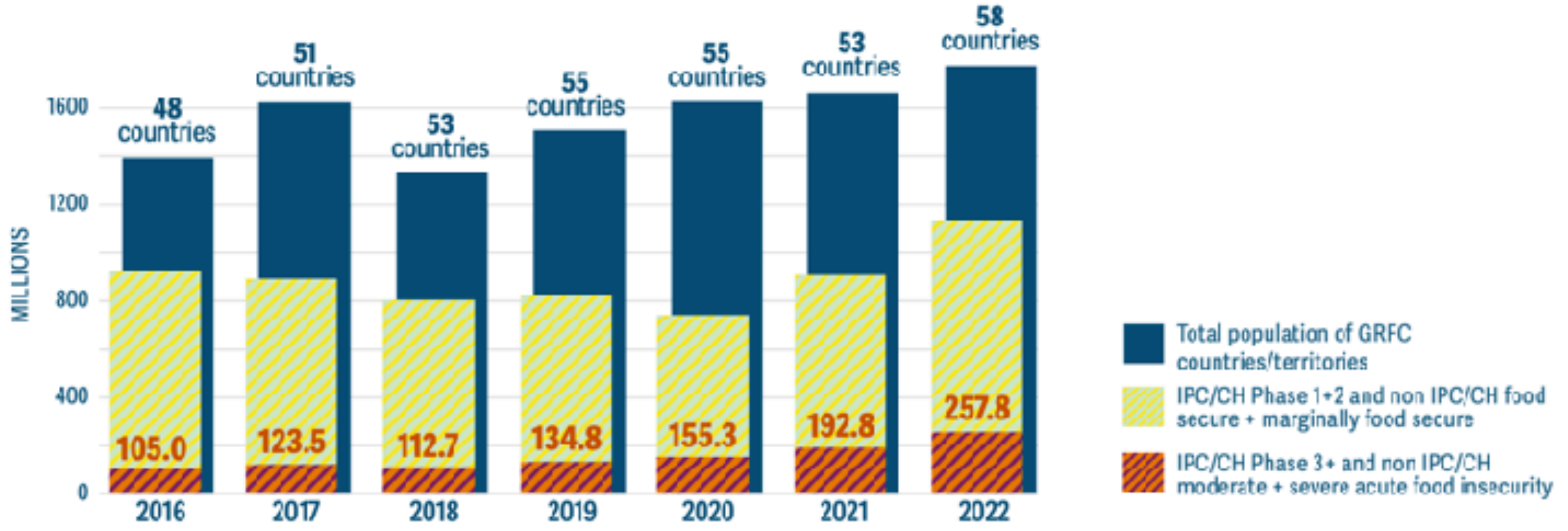


<0.5 million | 0.5–0.99 million | 1–2.99 million | 3–4.99 million | 5–9.99 million | 10–14.99 million | ≥15 million | Data not meeting GRFC requirements/population not analysed | Data gap | Country not selected for analysis | ○ Indicates migrants/refugee populations (colour coding as this key)

The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations. Final boundary between the Republic of Sudan and the Republic of South Sudan has not yet been determined. Final status of the Abyei area is not yet determined.

Source: FSIN, GRFC 2022.

Number of people in GRFC countries/territories facing acute food insecurity, 2016–2022



Source: FSIN, using data from 2016–2022.

IPC Global Initiative - Special Brief

IPC Special Briefs are produced by the IPC global initiative and do not necessarily reflect the views of stakeholders in Palestine. This analysis factors in all data and information available up to 10 March 2024 and does not take into account the latest developments on the ground.

THE GAZA STRIP

FAMINE IS IMMINENT AS 1.1 MILLION PEOPLE, HALF OF GAZA, EXPERIENCE CATASTROPHIC FOOD INSECURITY

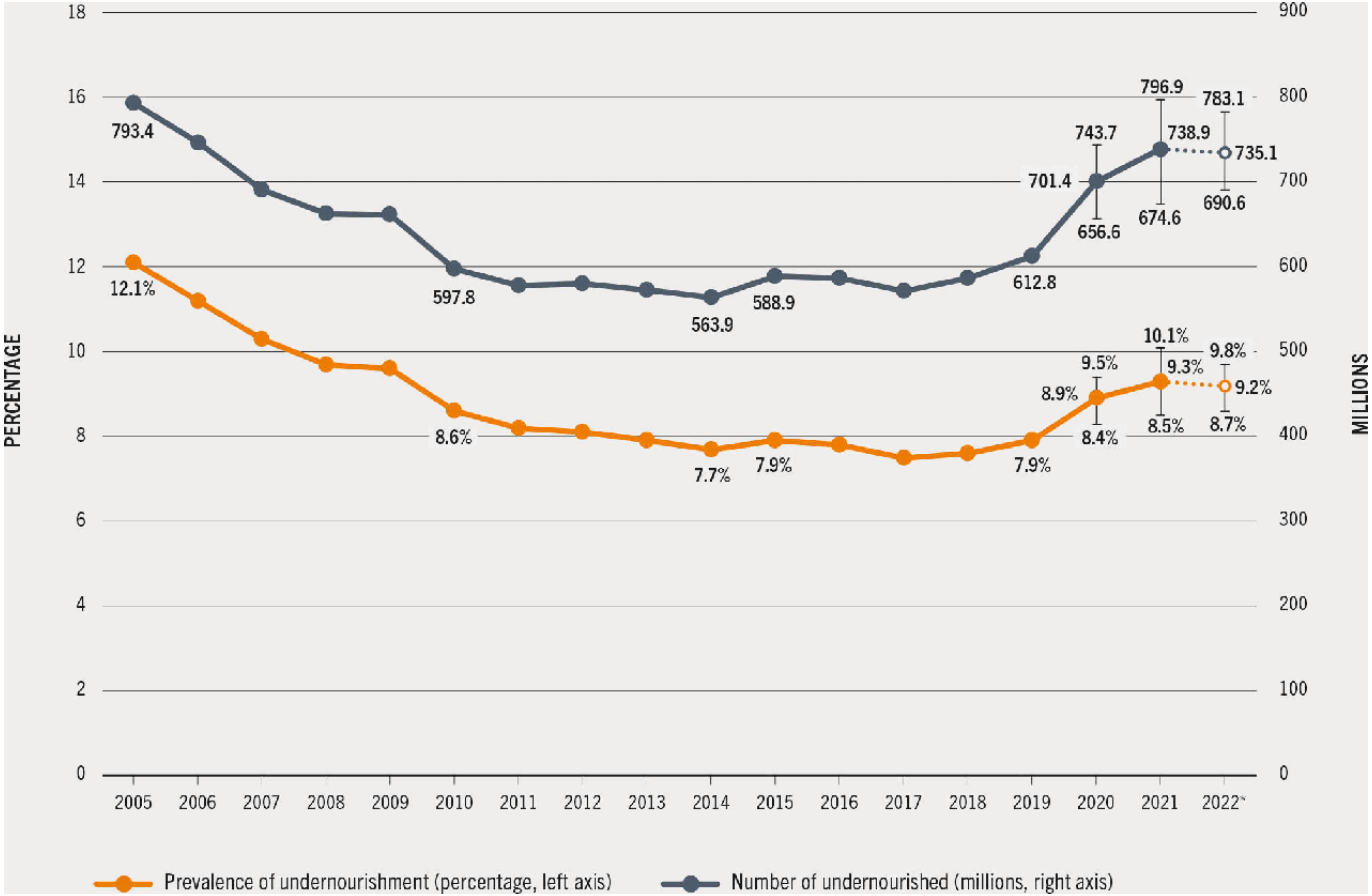
IPC ACUTE FOOD INSECURITY ANALYSIS
15 FEBRUARY - 15 JULY 2024
Published on 18 March 2024

CURRENT: 15 FEBRUARY - 15 MARCH 2024		
2.13M 95% of the population analysed	Phase 5	677,000 People in Catastrophe
People facing high acute food insecurity (IPC Phase 3 or above)	Phase 4	876,000 People in Emergency
	Phase 3	578,000 People in Crisis
IN NEED OF URGENT ACTION	Phase 2	96,000 People Stressed
	Phase 1	0 People in food security

PROJECTED: 16 MARCH - 15 JULY 2024		
2.23M 100% of the population analysed	Phase 5	1,107,000 People in Catastrophe
People facing high acute food insecurity (IPC Phase 3 or above)	Phase 4	854,000 People in Emergency
	Phase 3	265,000 People in Crisis
IN NEED OF URGENT ACTION	Phase 2	0 People Stressed
	Phase 1	0 People in food security

Global Report on Food Crises 2023

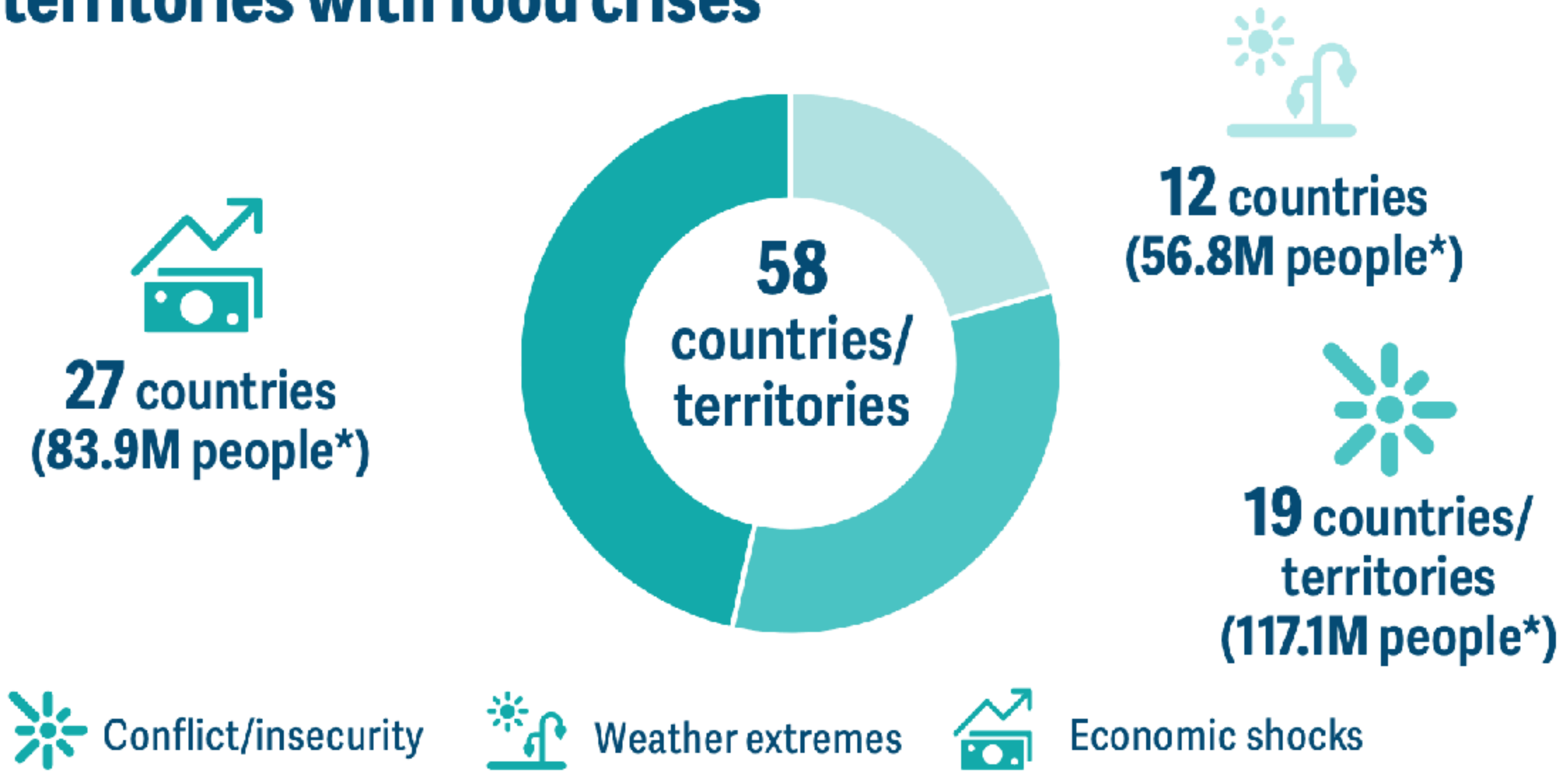
BETWEEN 691M AND 783M UNDERNOURISHED PEOPLE IN 2022



The State of Food Security and Nutrition in the World 2023

A COMPLEX PHENOMENON

Primary drivers of acute food insecurity in countries/ territories with food crises



Food crises are the result of multiple drivers. The GRFC has based this infographic on the predominant driver in each country/territory.

* Number of people in IPC/CH Phase 3 or above or equivalent.

Source: FSIN, GRFC 2023.

HOW DO WE MEASURE FOOD INSECURITY?

- **Food Consumption Score (FCS)**
- Household Dietary Diversity Score (HDDS)

} diversity of dietary intake

- **reduced Coping Strategies Index (rCSI)**
- Household Hunger Scale (HHS)

} consequences of constrained access to food

Vaitla et al. (2017) Food Policy 68

Integrated Food Security Phase Classification Technical Manual Version 3.0.

FOOD CONSUMPTION SCORE (FCS)

"In the last seven days, how often did you eat ______?"

- Main staples
- Pulses
- Fruit
- Vegetables
- Meat, fish, eggs
- Dairy products
- Sugar
- Oil

$$FCS = \sum w_i x_i$$

w_i = weight of food group *i*
x_i = frequency of consumption of *i*

Food consumption group	Standard threshold
Poor	0 – 21
Borderline	21 – 35
Acceptable	> 35

WFP (2008) Food consumption analysis

REDUCED COPING STRATEGIES INDEX (rCSI)

"In the last seven days, how often did you have to rely on one the following coping strategies?"

Coping strategy	**Severity weight**
Rely on less preferred or less expensive food	1
Borrow food or rely on help from friends or relatives	2
Limit portion size at mealtimes	1
Restrict consumption by adults in order for small children to eat	3
Reduce number of meals eaten in a day	1

$$rCSI = \sum w_i x_i$$

WFP (2008) The Coping Strategies Index: Field Methods Manual

DATA COLLECTION MODALITIES

- Face-to-face assessments
  - Comprehensive Food Security and Vulnerability Analyses (CFSVA)
  - Emergency food security assessments (EFSA)
- Mobile Vulnerability Analysis and Mapping (mVAM)
  - SMS
  - Interactive Voice Response (IVR)
  - Live calls

https://www.wfp.org/food-security-analysis

NEAR REAL-TIME FOOD SECURITY MONITORING

- WFP is currently conducting 2500+ household surveys *on a daily basis* through Computer Assisted Telephone Interviews (CATI) across 35 countries

Established Systems
35 countries

2018
- Nigeria
- Yemen
- Syrian Arab Republic

2019
West and Central Africa Expansion
- Burkina Faso
- Cameroon
- Central African Republic
- Chad
- Mali
- Mozambique
- Niger
- Democratic Republic of the Congo

January 2020
Central America Expansion
- Colombia
- El Salvador
- Guatemala
- Honduras

March 2020
- Afghanistan
- Iraq

May - August 2020
May
- Haiti
- Malawi

June
- Côte d'Ivoire
- Ethiopia
- Madagascar
- United Republic of Tanzania
- Sierra Leone

July
- Republic of the Congo

August
- Somalia
- Zambia
- Nicaragua

September - November 2020
September
- Guinea
- Zimbabwe

October
- Angola
- Kenya

November
- Benin

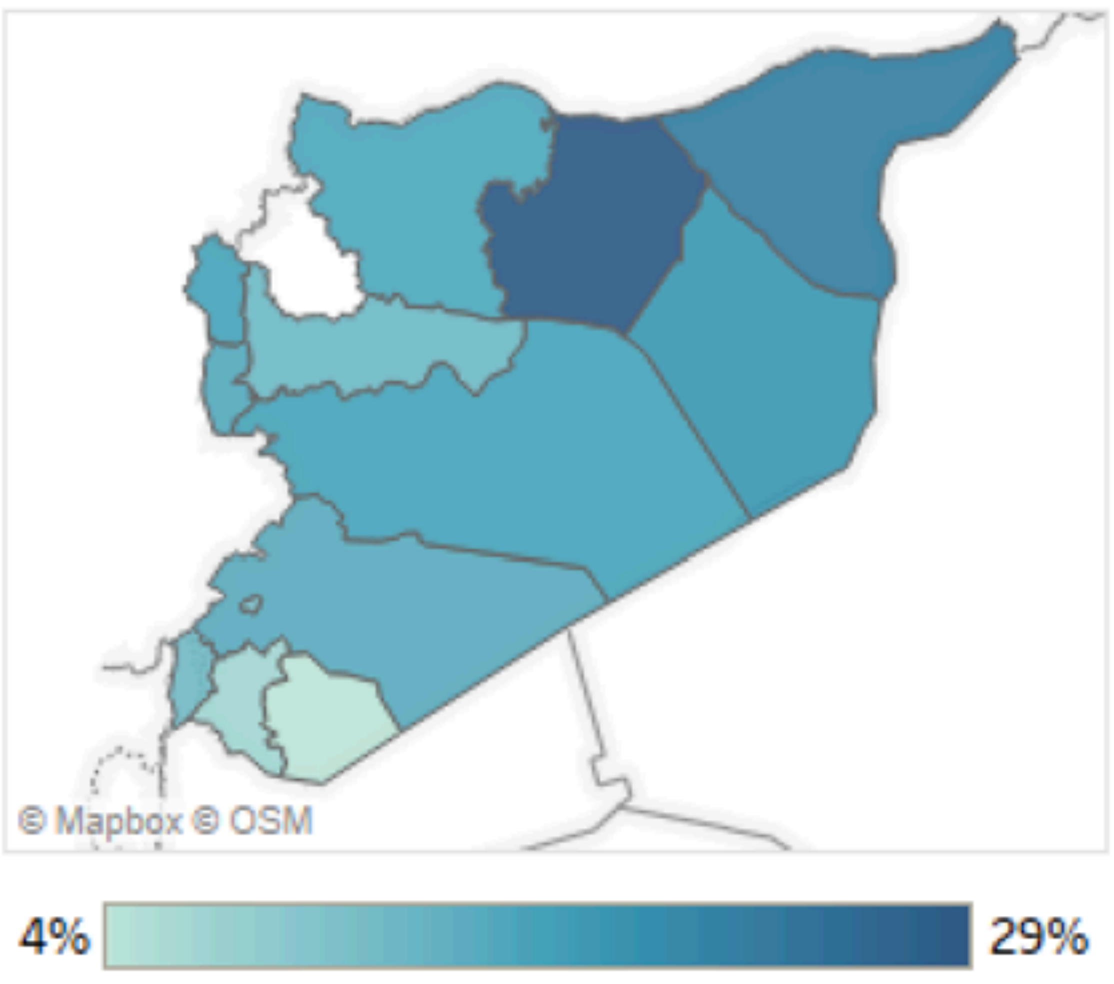
December 2020-Mid 2021
- Mauritania
- Namibia

FROM HOUSEHOLD LEVEL INFORMATION TO REGIONAL CHARACTERIZATION

- Prevalence of households with poor or borderline food consumption (FCS <= 35)
  –> **Prevalence of people with insufficient food consumption**

- Prevalence of households with rCSI >= 19
  –> **Prevalence of people using crisis or above crisis food-based coping**

Map1: Percent of households by governorate reporting poor food consumption in June 2021

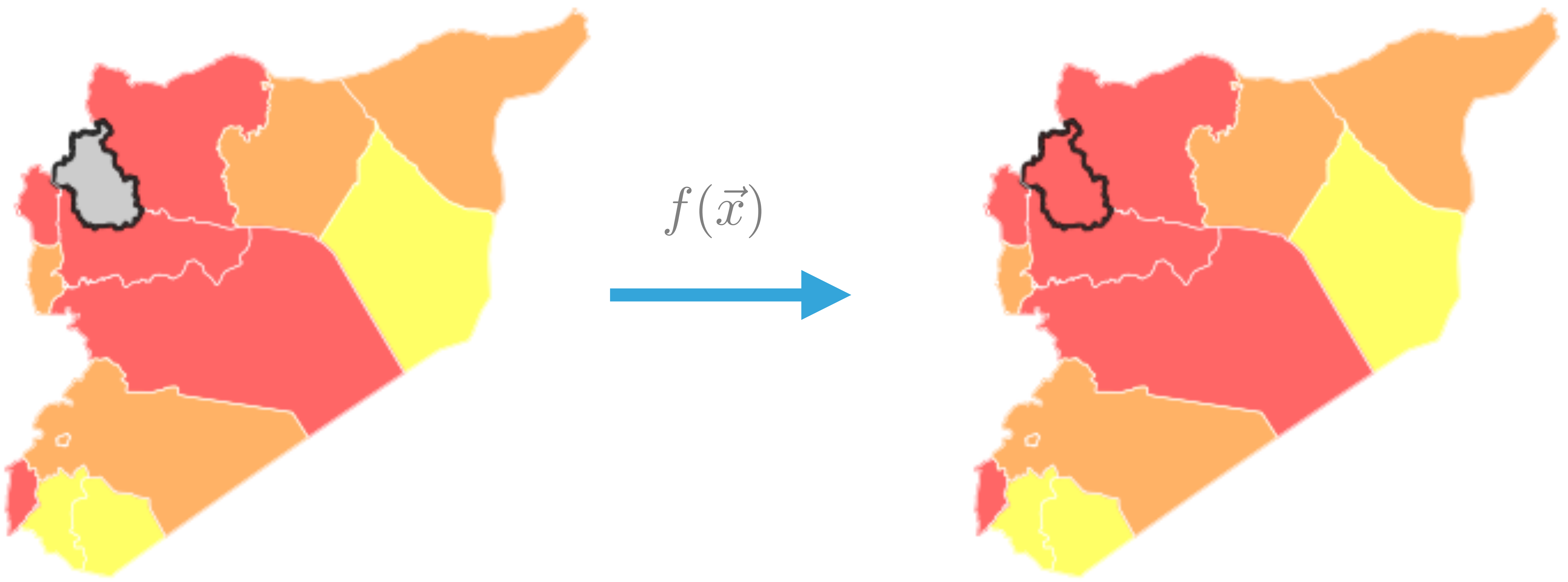


Data from Idleb is not available

Syria mVAM Bulletin Issue no. 56

RESEARCH QUESTION

- Can we estimate the current food security situation in a given area when no primary data is available?



MAIN DRIVERS OF FOOD INSECURITY

Conflict/insecurity

Weather extremes

Economic shocks

COMBINING DIVERSE DATA SOURCES

- Conflict information (ACLED)
- Market prices (WFP)
- Macro-economic indicators (Trading Economics)
- Undernourishment (FAO/WFP)
- Rainfall (CHIRPS)
- Vegetation Index (MODIS)

COMBINING DIVERSE DATA SOURCES

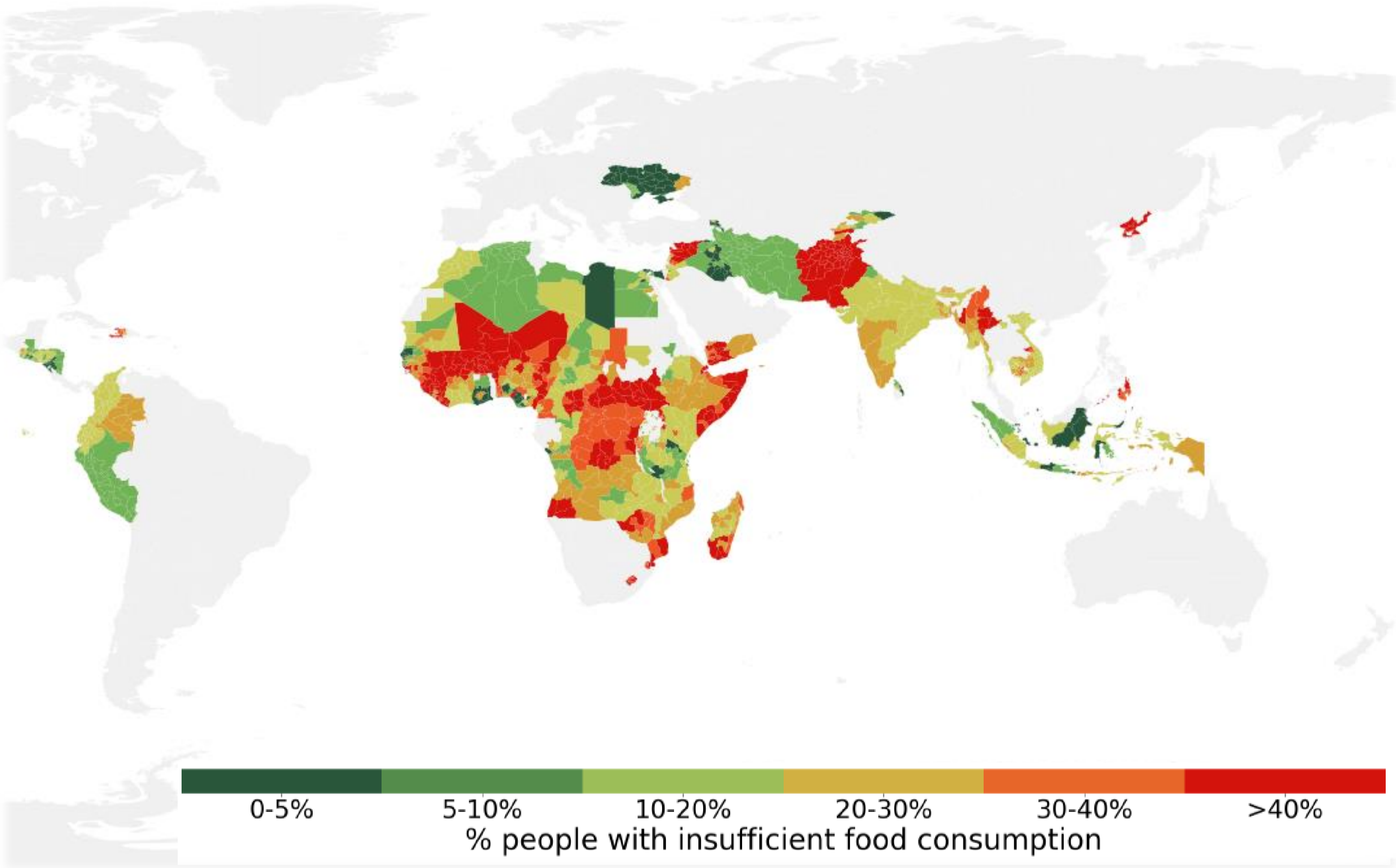
- Conflict information (ACLED)
- Market prices (WFP)
- Macro-economic indicators (Trading Economics)
- Undernourishment (FAO/WFP)
- Rainfall (CHIRPS)
- Vegetation Index (MODIS)

COMBINING DIVERSE DATA SOURCES

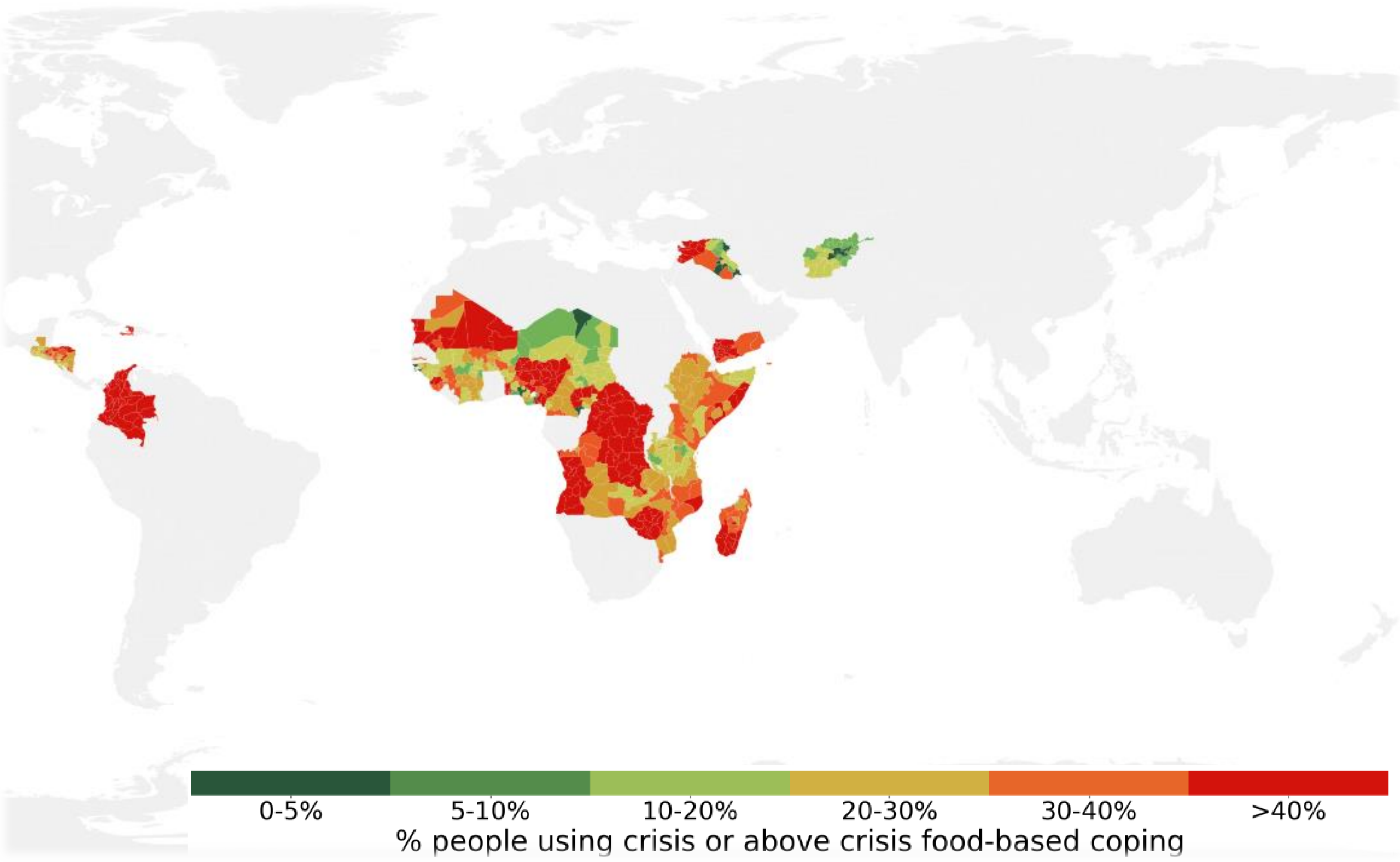
- Conflict information (ACLED)
- Market prices (WFP)
- Macro-economic indicators (Trading Economics)
- Undernourishment (FAO/WFP)
- Rainfall (CHIRPS)
- Vegetation Index (MODIS)

TRAINING DATA

- **Prevalence of people with insufficient food consumption**
  ~38K observations across 78 countries and 15 years (~24K with previous assessment)



- **Prevalence of people using crisis or above crisis food-based coping**
  ~36K observations across 41 countries and 8 years (~12K with previous assessment)



$\vec{x}(i,t)$

Previous food consumption assessment (when available)

Last available undernourishment estimate

Increase/decrease in conflict fatalities

Rainfall & its anomaly

Vegetation index (NDVI) & its anomaly

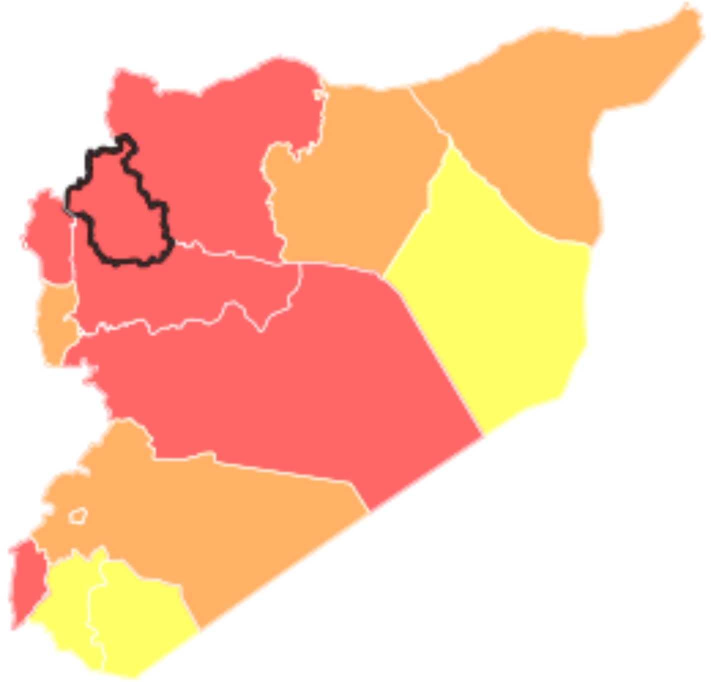
Cereal price variation

Headline & food inflation

Exchange rate % change

$y(i,t)$

% people with insufficient food consumption in area *i* at time *t*

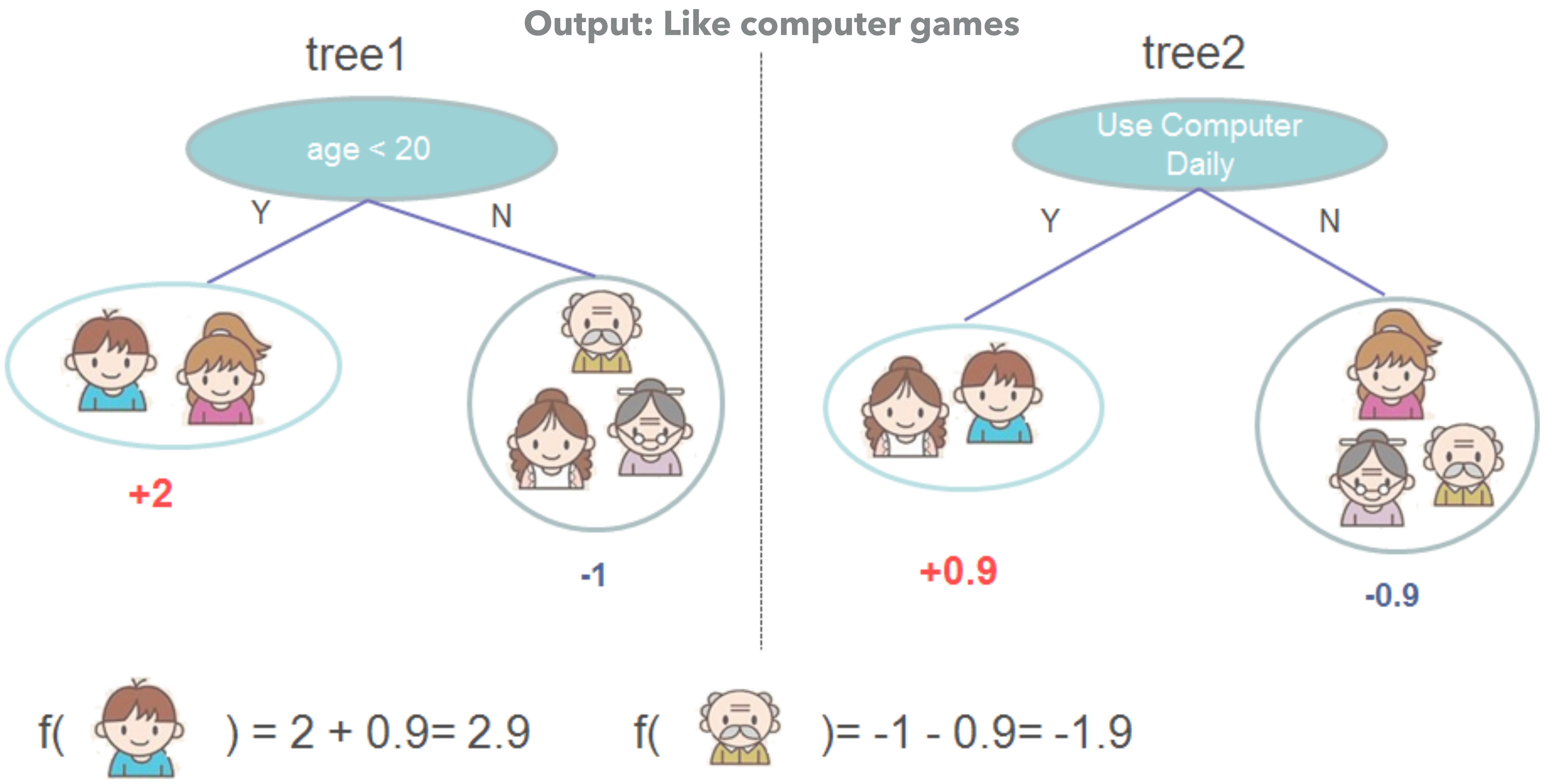


FOUR MODELS (TWO PER TARGET VARIABLE)

- **Prevalence of people with insufficient food consumption**
  - Previous prevalence included as input variable
  - Previous prevalence **not** included as input variable
- **Prevalence of people using crisis or above crisis food-based coping**
  - Previous prevalence included as input variable
  - Previous prevalence **not** included as input variable

MODELING APPROACH

- Gradient boosted decision tree ensembles (**XGBoost**, https://xgboost.readthedocs.io/en/stable/)



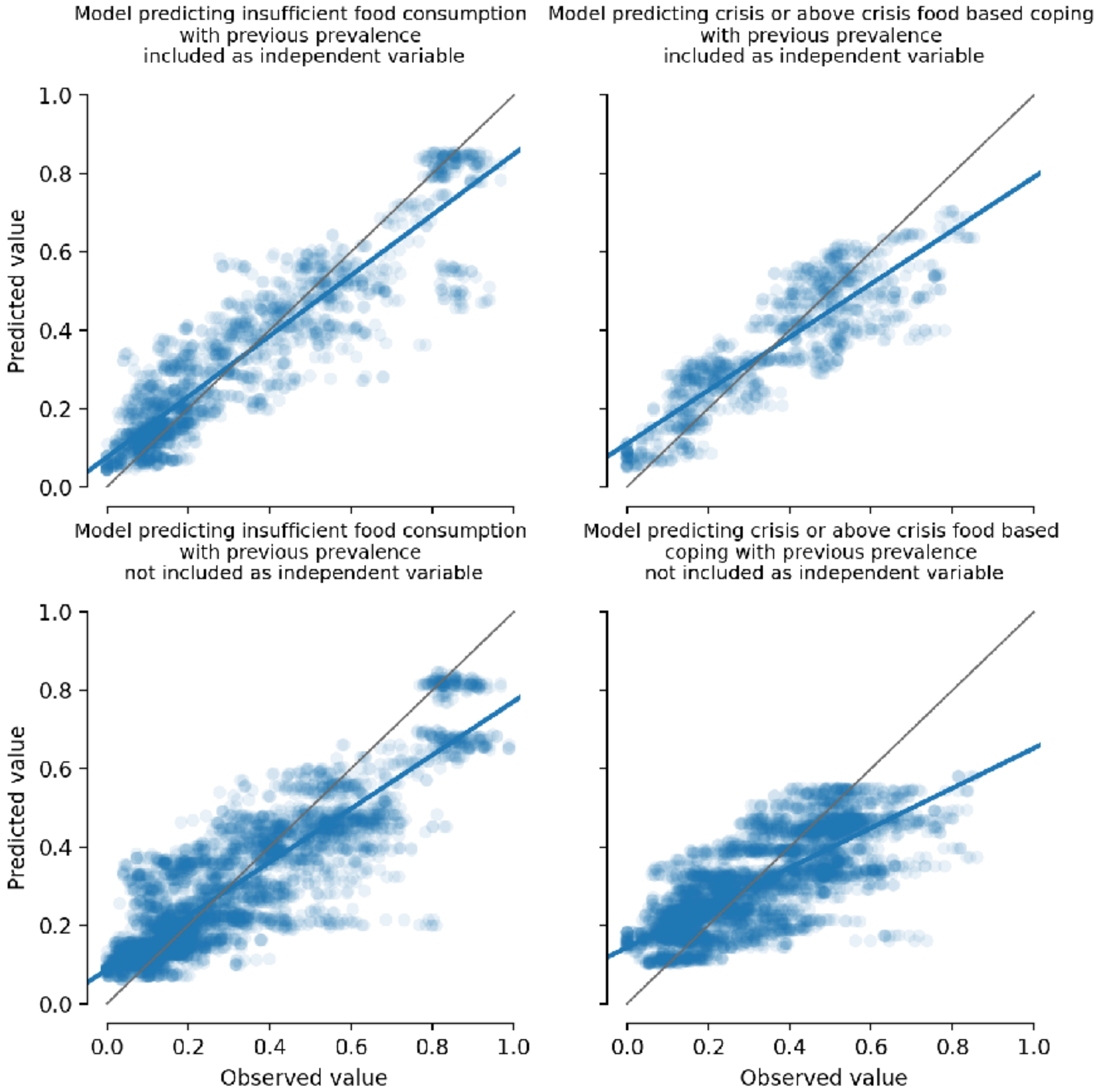
T. Chen, Introduction to Boosted Trees (2014)

TRAINING & VALIDATION APPROACH

- **Temporal** split:
  - **Training & validation**: data until 31 May 2021 (~85% of the data)
  - **Testing**: the remaining two months (~15% of the data)
- Model **hyper parameters tuning** with a **walk-forward validation** approach:
  4 folds were created (each covering 1 month of data) from February through May 2021, and for each fold, the training set was composed of all the older data up to the end of the previous month
- The chosen combination of hyper parameters is the one leading to the **smallest difference** between the average **R^2** on the folds used as training set and the average R^2 on the folds used as validation (to avoid overfitting)
- Once the hyper parameters are selected, N_b = 100 models are fitted on samples with replacement of the training and validation set (**bootstrapping**), and the test set is used to evaluate the model's performance
- Final prediction are given by the **median of the N_b predictions + 95% CI**

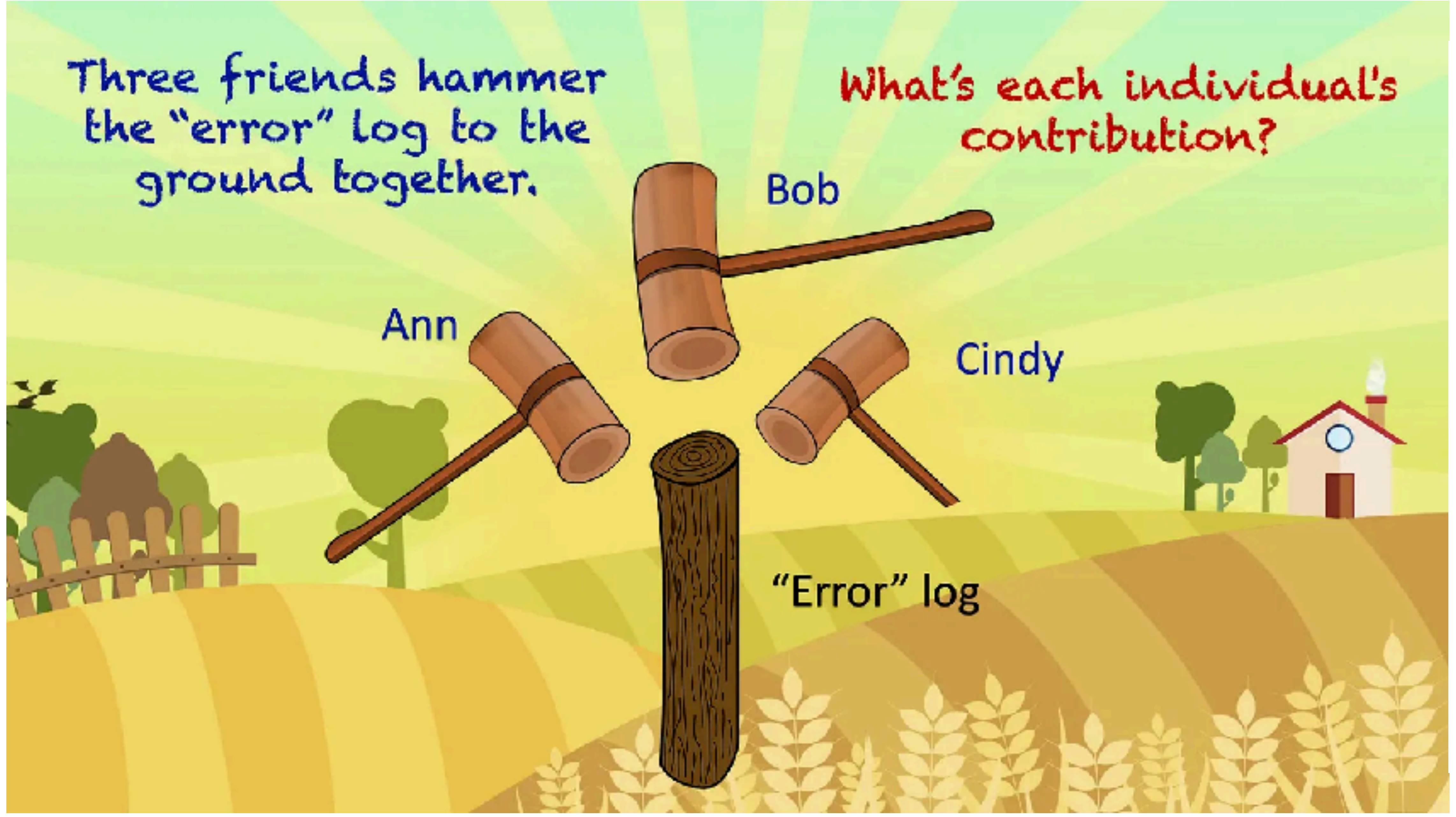
RESULTS

		R^2	MAE
Food consumption	Prevalence from previous assessment included as independent variable	0.81	0.07
	Prevalence from previous assessment **not** included as independent variable	0.74	0.09
	Naive model	0.51	0.15
Food-based coping	Prevalence from previous assessment included as independent variable	0.73	0.08
	Prevalence from previous assessment **not** included as independent variable	0.61	0.09
	Naive model	0.45	0.12

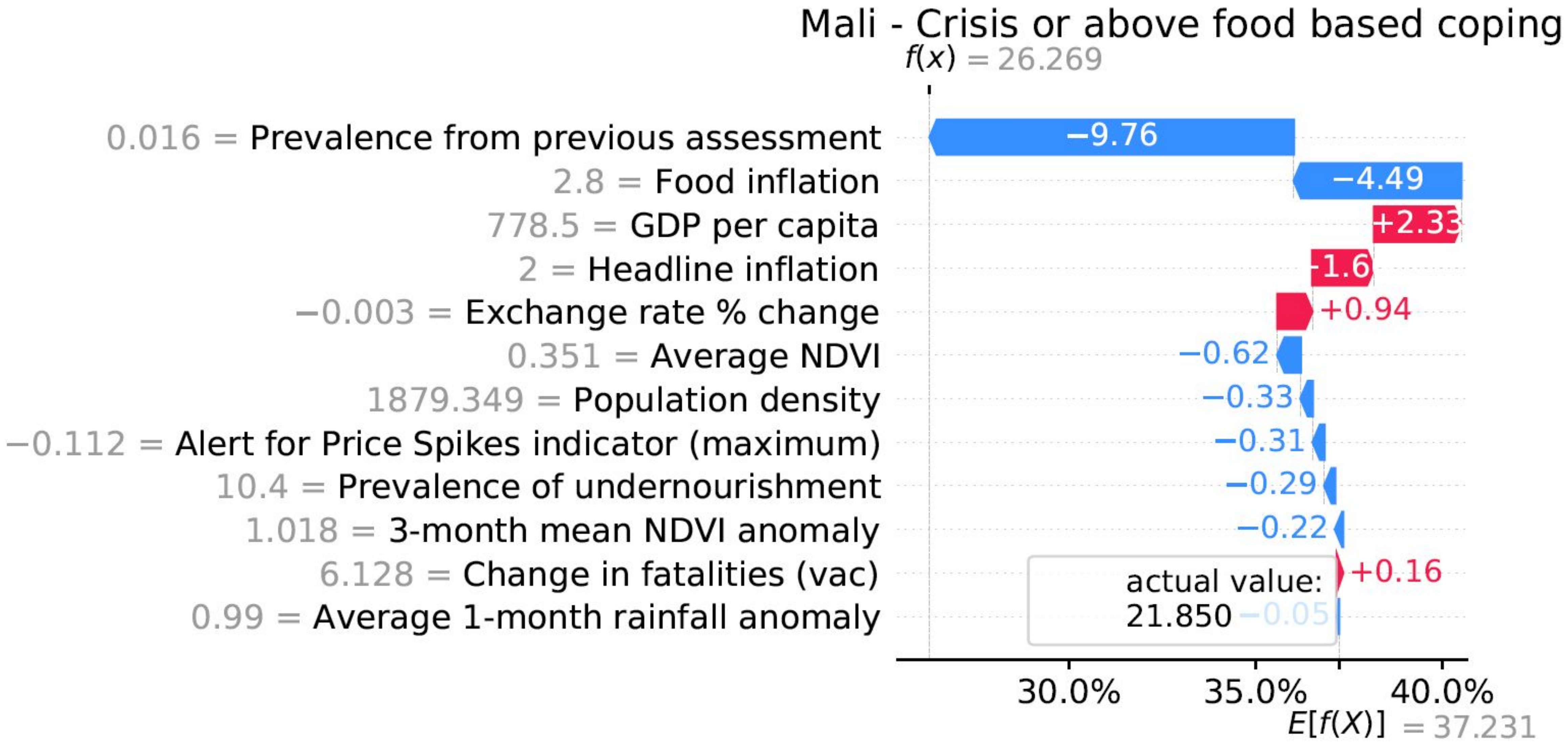


EXPLAINING PREDICTIONS

- **SHAP values** (SHapley Additive exPlanations, https://shap.readthedocs.io/en/latest/) can provide an easily interpretable explanation of individual predictions

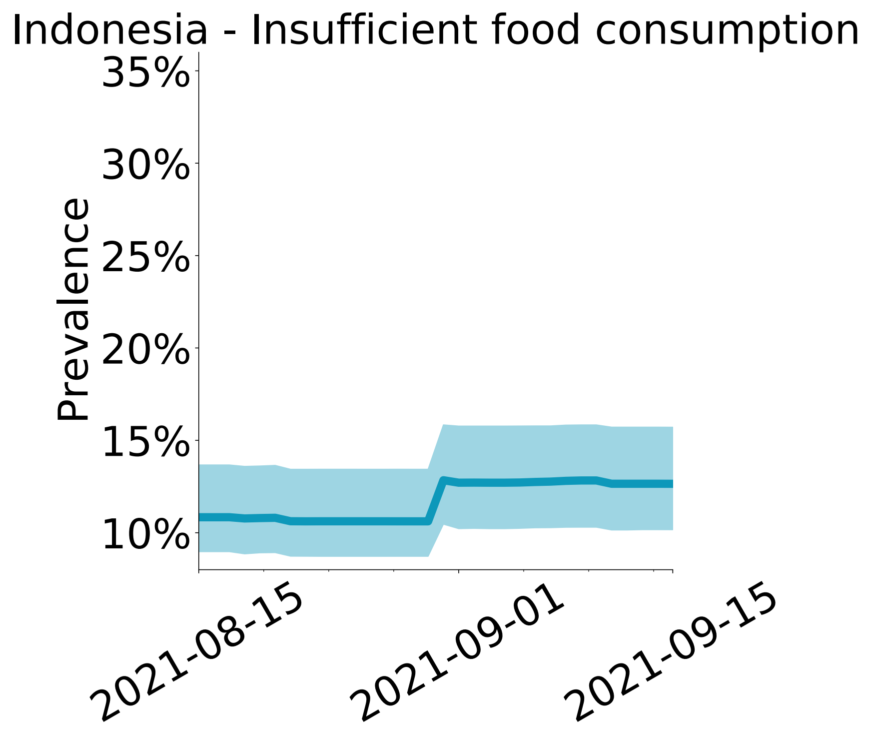


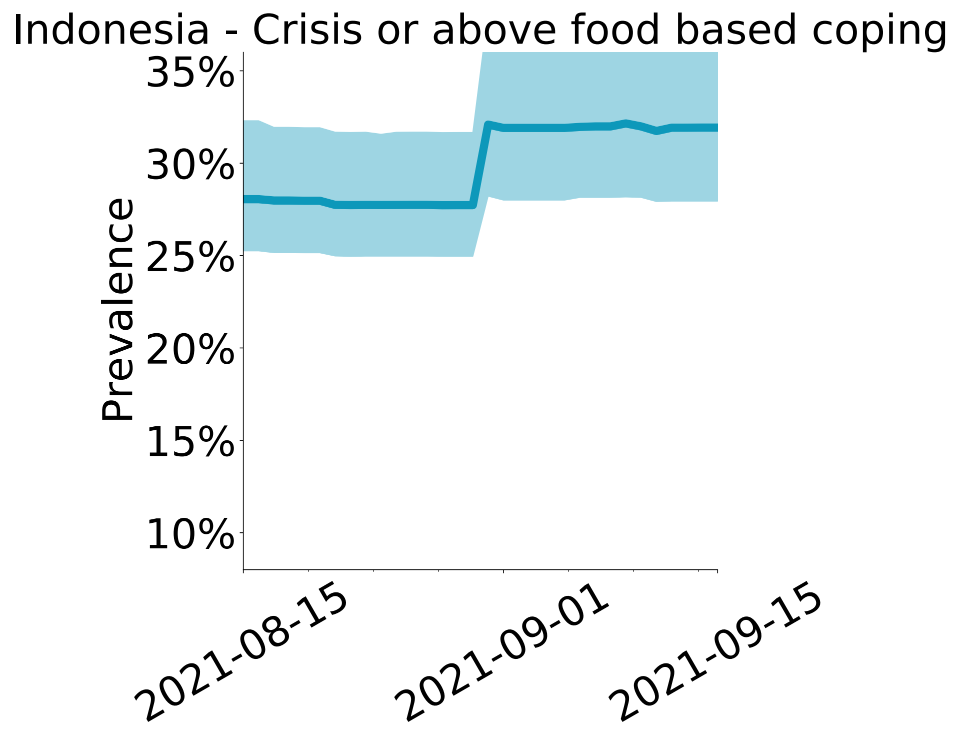
https://medium.com/dataman-in-ai/explain-your-model-with-the-shap-values-bc36aac4de3d

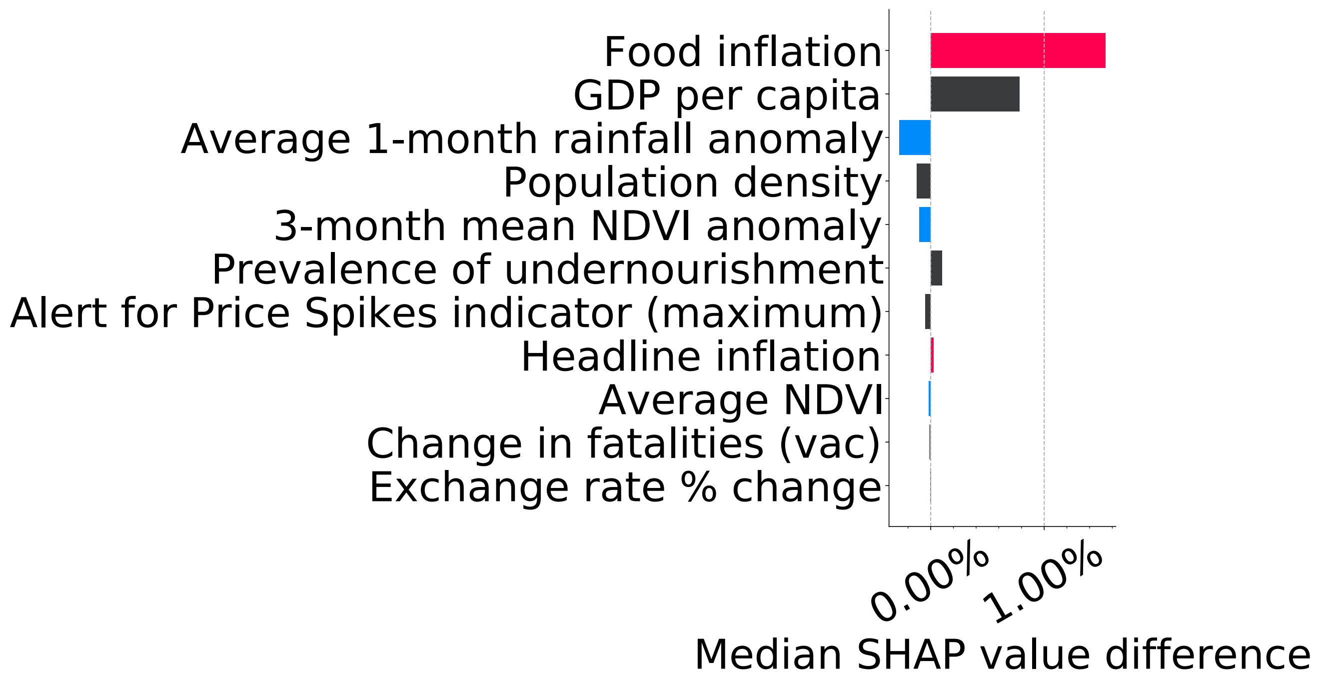


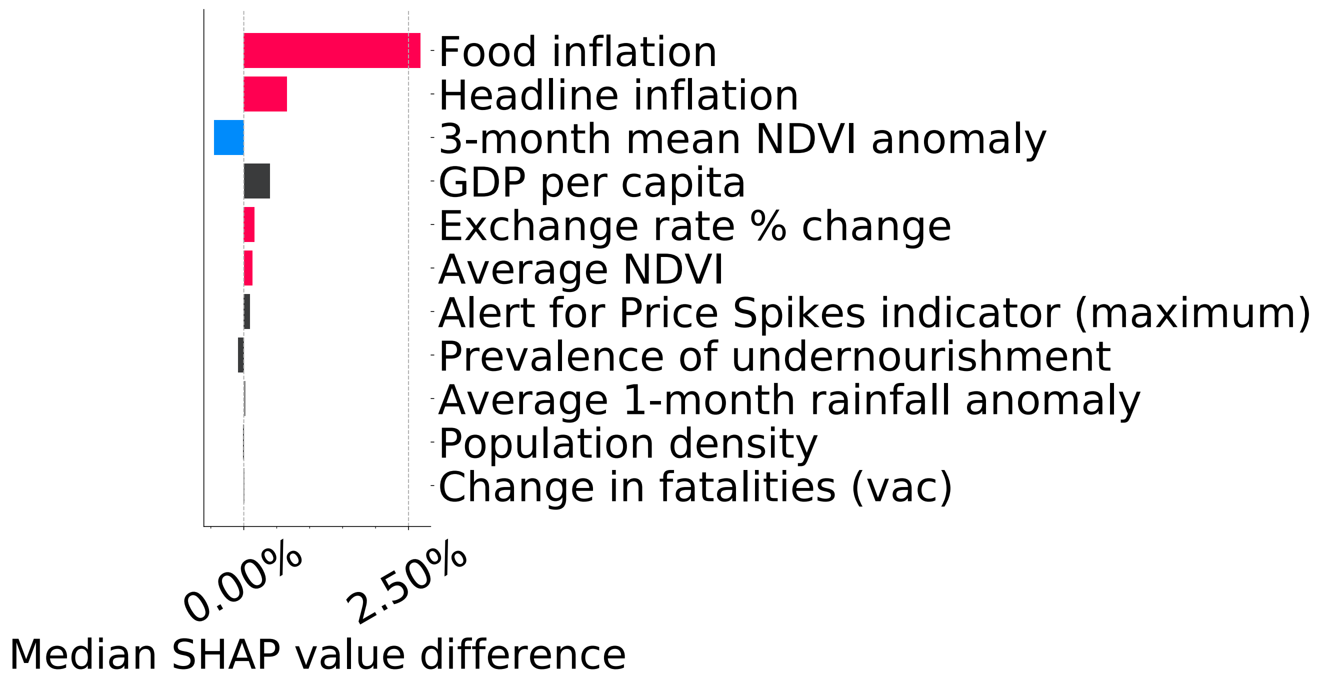
EXPLAINING PREDICTIONS AND TREND CHANGES

- SHAP value differences can explain deterioration/improvement of the situation over time









	2021-08-15	2021-09-15
Food inflation	2.740	3.310
Average 1-month rainfall anomaly	125.779	147.621
3-month mean NDVI anomaly	1.022	1.039
Headline inflation	1.520	1.590
Average NDVI	0.749	0.749
Change in fatalities (vac)	0.022	0.047
Exchange rate % change	0.078	0.120

	2021-08-15	2021-09-15
Food inflation	2.740	3.310
Headline inflation	1.520	1.590
3-month mean NDVI anomaly	1.022	1.039
Exchange rate % change	0.078	0.120
Average NDVI	0.749	0.749
Average 1-month rainfall anomaly	125.779	147.621
Change in fatalities (vac)	0.022	0.047

MODEL OPERATIONALIZATION

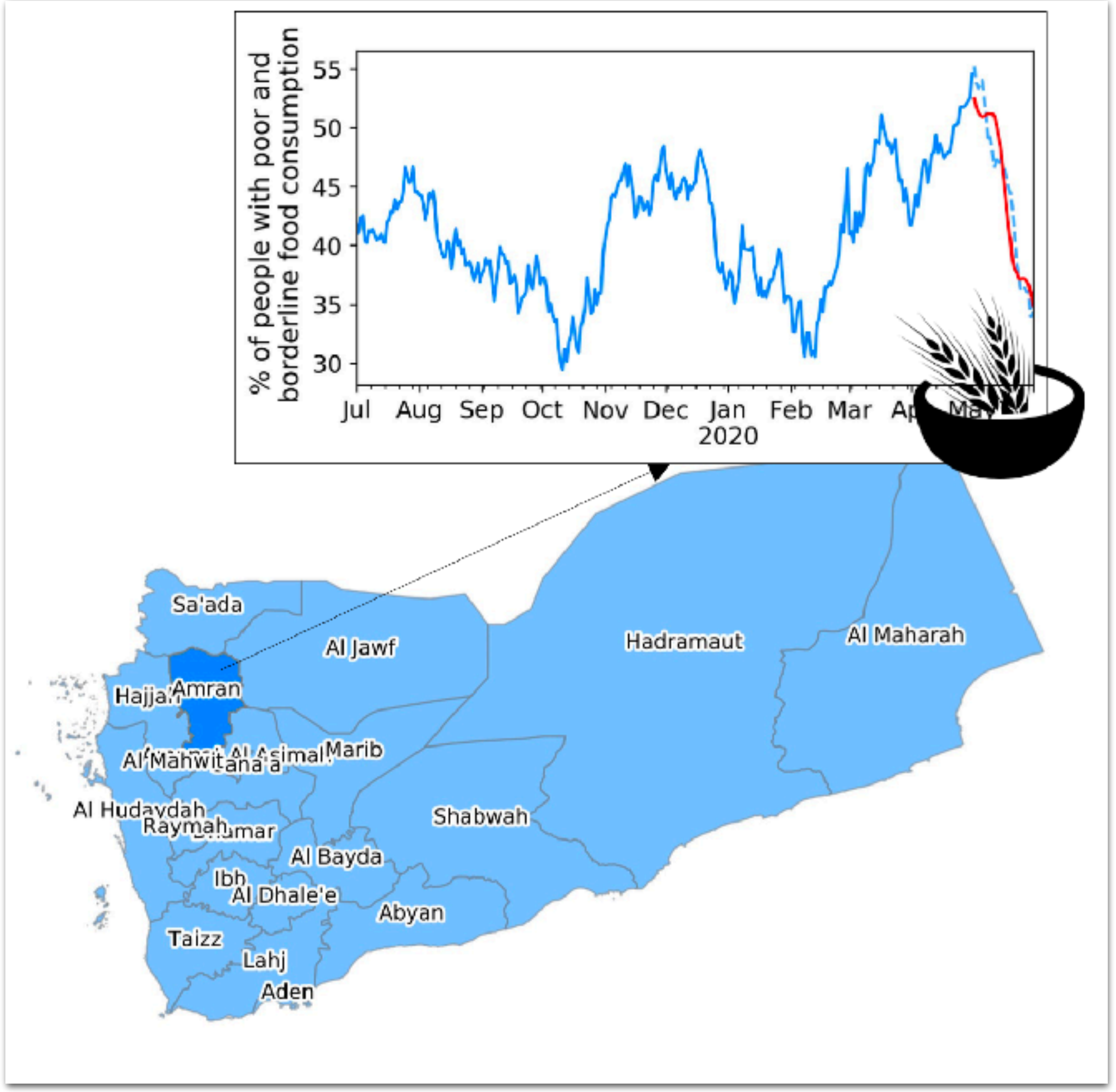
hungermap.wfp.org

CHALLENGES & LIMITATIONS

- Global model vs local approach
  - Sampling of training data to account for unequal geographical representation
  - Smaller sensitivity to local patterns than local models
  - Transferability to geographic contexts not included in the training data is not guaranteed
- Open data availability on a global scale at the same geographical and temporal resolution is a challenge
- Predictions should be handled with caution and used only to trigger further assessments of the situation

FROM NOWCASTING TO FORECASTING

- Where food security primary data is available, can we forecast its evolution in the near future?



THE ROLE OF CONFLICT

- Do all conflict events affect food security in the same way?

Yemen	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$R2 = 0.32$	$R2 = 0.32$	$R2 = 0.32$	$R2 = 0.32$	$R2 = 0.32$	$R2 = 0.32$
Variable	**Coefficients**					
const	42.812***	42.545 ***	43.343***	42.453***	42.537***	42.739***
FRConflicts	0.257***					
NFRConflicts		0.006***				
VaC			-0.056***			
Battles				0.008***		
Total Conflicts					0.006***	
Past Conflicts						0.004***
Pop	-6.996 ***	-6.877 ***	-6.639***	-6.858***	-6.881***	-6.863***
Rainfall	-0.047***	-0.065***	-0.070***	-0.069***	-0.064***	-0.058***
NDVI	76.312***	76.710***	74.905***	76.979***	76.738***	76.070***
ALPS	0.145	0.161	0.025	0.166	0.162**	0.102

*** pval < 0.001, ** pval < 0.01, * pval < 0.05

COLLABORATORS

- UN World Food Programme: Alberto Bracci, Matteo Corea, Arif Husain, Sejal Jaiswal , Giulia Martini, Lorenzo Riches, Jonathan Rivers
- ISI Foundation: Stefania Fiandrino, Pietro Foini, Yelena Mejova, Daniela Paolotti, Michele Tizzani, Michele Tizzoni
- Dublin City University: Caitriona Dowd

PUBLICATIONS

G. Martini, A. Bracci, L. Riches, S. Jaiswal, M. Corea, J. Rivers, A. Husain, **E. Omodei**.
Machine learning can guide food security efforts when primary data are not available.
Nature Food 3, 716–728 (2022)

P. Foini, M. Tizzoni, D. Paolotti, **E. Omodei**.
On the forecastability of food insecurity.
Scientific Reports 13, 2793 (2023)

S. Fiandrino, C. Dowd, G. Martini, Y. Mejova, **E. Omodei**, D. Paolotti, M. Tizzani.
Impact of food-related conflicts on self-reported food insecurity.
Frontiers in Sustainable Food Systems, Volume 7 (2023)

E. Omodei. Using computational tools to monitor and improve access to quality food and water.
Nature Computational Science 3, 726–728 (2023)

CAN YOU THINK OF EXAMPLES WHERE DATA SCIENCE CAN INSTEAD INHIBIT THE ACHIEVEMENT OF THE SDGS?

- Societal biases in training data will inadvertently learn and reproduce these biases
  - Lack of diversity in image recognition training sets
  - Gender stereotypes in language
  - Racial and socioeconomic biases in criminal justice
- Lack of gender, racial, and ethnic diversity also in the AI workforce
- Personalized content on social media may lead to political polarization and affect social cohesion
- High energy needs for AI applications might hinder efforts to achieve climate action

Vinuesa et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020)

OUTLOOK

Main challenges in the use of digital data and computational methods to monitor the SDGs

Omodei *et al.* Journal of Physics: Complexity (2022)

BIG DATA UNDER-REPRESENTS THE MOST VULNERABLE

- Digital data often more available and more granular than traditional data
- However, not designed and collected to answer a research question
- Not equally produced by everybody and under-representing disadvantaged geopolitical and socioeconomic contexts

LACK OF GOOD DATA-SHARING PARADIGMS

- Issues such as data privacy, governance, and responsible use
- Main challenge to preserve users’ privacy through anonymization & aggregation of row data while loosing as little information as possible
- Long & time consuming processes that often hinder the development and applicability of models during emergencies

REPLICABILITY AND TRANSFERABILITY TO LOW INCOME SETTINGS

- The majority of available studies focuses on big cities and (data) rich countries
- Transferability to vulnerable and data-poor contexts is not guaranteed
- Validity over time is also not guaranteed

OPERATIONALIZATION

- Model’s interpretability and explainability
- Transitioning research efforts into operational near real-time tools
- Integration into existing systems and scale-up

CEU | CENTRAL EUROPEAN UNIVERSITY

Thank you!

omodeie@ceu.edu | @elisa_omodei