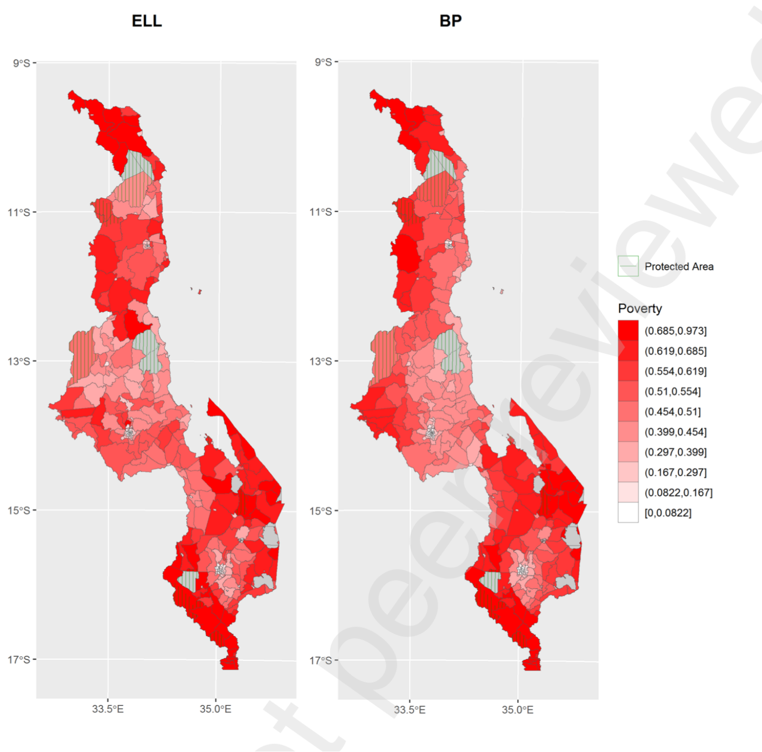
MERFELD AND NEWHOUSE

- Applied in Mexico
- Main analysis using monetary poverty estimates at the household level from the 2014 MCS-ENIGH Household Survey data.
- Two CNNs made.The primary input to the CNN models is high-resolution satellite imagery. This imagery has spatial pixel resolution ranging from 3 to 5 metres. It captures visual characteristics of the areas under consideration.
    - The CNN for poverty prediction generates predictions related to <u>extreme or moderate headcount poverty</u>.
    - The CNN for land classification generates predictions for the type of land present in the image tile, such as whether it's a building, road, water, grassland, forest, or background.
- Three levels created- household level, sub-area level and area level poverty.
- Accuracy of model quite high, especially at the household-level

VAN DER WEIDE

- It first obtains small area estimates of poverty by combining household expenditure survey data with population census data. It then ignores the population census and obtains a second poverty map by combin- ing the survey with predictors of poverty derived from remote sensing data.
- The research relies on the 2010-11 Malawi Third Integrated Household Survey (IHS3) for household consumption expenditure data to measure poverty.
- The poverty line is defined as the percentage of individuals whose annual household consumption per capita falls below the national poverty line, which is expressed in Malawi kwacha.
    - ELL APPROACH-The study combines the IHS3 data with the 2008 Population and Housing Census data to estimate poverty at the small area level. The census provides information on household composition, education, employment, dwelling characteristics, and asset ownership. This information is used to impute household consumption expenditures, which are then aggregated for poverty estimation.
    - SEM APPROACH- the IHS3 data is combined with remote sensing data like <u>night-time lights, urban footprints, major roads, population density, vegetation, surface temperature, and rainfall.</u> These data are used to impute village-level poverty rates.
- Study aimed at creating poverty maps.
- Correlation between the geospatial features and census estimates above 0.9.



MASAKI ET AL.

- Applied in Tanzania and Sri Lanka
- Found that combining survey data with geospatial data significantly improves both the precision and accuracy of our non-monetary poverty estimates
- Non-monetary poverty is directly observed in census data. Household assets and demographic proxies are identified to measure welfare. Principal component analysis is used to estimate loading factors for these indicators.
- Synthetic surveys are created by merging actual household budget survey data with census data at the village level.
- Geospatial data are drawn from various publicly available sources. These include night-time lights, precipitation, elevation, forest cover change, built-up area, population estimates, climate classification, and crop yield estimates. For Sri Lanka, contextual features derived from Sentinel-2 imagery are also used to capture correlations with poverty, population density, buildings, and roads. Proprietary building footprint data from Ecopia and Maxar are used for Tanzania.
- The model produces estimates of non-monetary poverty rates at small area levels. These estimates are derived using various statistical techniques, including the Fay-Herriot area-level model and the household-level model.
- **Fay-Herriot Area-Level Model:** This model is used to estimate small area parameters based on area-level aggregated data. It's named after its creators Fay and Herriot. It's a form of empirical best linear unbiased prediction (EBLUP) and is a form of linear regression that includes a small area random effect.

- **Household-Level Model:** This model, also referred to as the EBP (Empirical Best Predictor) model, operates at the individual household level. It takes into account individual-level survey data as well as area-level geospatial and auxiliary data.
- **The correlation with the census rose from 0.72 to 0.88 in Sri Lanka and from 0.77 to 0.88 in Tanzania.**

KRENNMAIR AND SCHMID

- This semi-parametric approach is promising as it avoids the assumptions of linear mixed models in contrast to classical small area models and builds on random forests. These tree-based machine learning predictors have the advantage of robustness against outliers and implicit model-selection.
- The input data consists of socioeconomic and demographic information from households in Mexican municipalities, specifically from the state of Nuevo León.
- The dataset includes variables related to household income, socio-demographic factors, educational levels, occupation, sources of income, etc.
- The input data is derived from two sources: the Mexican household income and expenditure survey (ENIGH) and census microdata by the National Institute of Statistics and Geography.
- The main output is the estimation of average household income per capita (ictpc) for different municipalities in the state of Nuevo León.
- The primary machine learning model used in this application is the Mixed-effects Random Forests (MERF) method.
- The MERF model is compared with other methods including EBP-BC, BHF, and P-SPLINES.