FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

HARNESSING BIG DATA AND MACHINE LEARNING TO FEED THE FUTURE

WEBINAR 2

Using Big Data and Machine Learning to Predict Poverty and Malnutrition for Targeting, Mapping, Monitoring, and Early Warning

July 28, 2021 / 8:00 AM – 9:30 AM EDT

PRESENTERS

Linden McBride
Assistant Professor, St. Mary's College of Maryland

Yanyan Liu
Senior Research Fellow, Markets, Trade and Institutions Division, IFPRI, USA

Chris Browne
PhD Candidate, Center for Applied Mathematics, Cornell University, USA

Medha Bulumulla
Research Assistant, Cornell University, USA

MODERATOR

Christopher B. Barrett
SB&JG Ashley Professor, Cornell University, USA

DISCUSSANT

Rob Vos
Division Director, Markets Trade and Institutions Division, IFPRI, USA

USAID FROM THE AMERICAN PEOPLE

FOOD SECURITY PORTAL FACILITATED BY IFPRI

SUPPORTED BY European Commission

Cornell University

CGIAR RESEARCH PROGRAM ON Policies, Institutions, and Markets

Led by IFPRI

Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning

Applied Economics Perspectives & Policy

IFPRI, July 28 2021

Linden McBride, Christopher B. Barrett, Christopher Browne, Leiqiu Hu, Yanyan Liu, David Matteson, Ying Sun, and Jiaming Wen

Trade offs, innovations, and frontiers

Can big data revolutionize poverty and malnutrition mapping, targeting, M&E and forecasting?

Agencies' needs vary; consider purpose, use case, of the tool

- What type of deprivation is being mapped/targeted/monitored/forecasted?
- What is the time horizon?
- How transparent/accessible does the final model need to be?
- How onerous is the data collection and curation task?
- What is the agency's objective function (Zhou et al. 2021)?

Fit tools to tasks

- **Targeting versus mapping**: Targeting identifies poor/malnourished people while mapping identifies poor/malnourished places

- **Structural versus stochastic**: Tension between asset based theory, and empirics, of poverty traps and big data/ML

- **Static versus dynamic**: Mapping and targeting efforts tend to produce static models whereas early warning systems identify those who will be poor/malnourished/food insecure in the next period/anticipate the impact of shocks on vulnerable populations

- **Data needs**: ML informed map, tool, or model will only be as good as the data on which it is trained and tested

Targeting innovations

Scorecard approach to proxy means test development using machine learning for dimension reduction and out of sample validation for model assessment

- Lean data (Schriener 2007, Kshirsagar et al. 2017, Baez et al. 2019)
- High frequency data (Knippenberg et al. 2019)
- Administrative data (Altındağ et al. 2021)

Mapping innovations

Combining different data inputs using CNN or other ML methods to estimate measures of deprivation at local levels

- Cell Data Records (Blumenstock et al. 2015)
- Nightlights, daytime satellite imagery, NDVI, remotely sensed data (Jean et al. 2016, Yeh et al. 2020)
- Combined data sources (Pokhriyal & Jacques 2017, Yeh et al. 2020)
- Multidimensional Poverty Index (Pokhriyal & Jacques 2017, Njuguna & McSharry 2017)
- Open source data (Hersh et al. 2020)
- Multivariate prediction of correlated outcomes (Browne et al. 2021)

Targeting and mapping frontiers

- Multi-dimensional nature of deprivation
- Predicting low probability/noisy outcomes (Head et al. 2017)
- Limitations of available data (Blumenstock 2016 & 2020)
- Determinants of geographically concentrated poverty (Yeh et al. 2020)

Structural versus stochastic

Innovations

- Recent mapping models do relatively well predicting asset poverty across space (Blumenstock et al. 2015, Jean et al. 2016, Yeh et al. 2020, Browne et al. 2021) and over time (Yeh et al. 2020, Browne et al. 2021)

Frontiers

- Parsing persistently poor from the dynamically mobile (Carter & Barrett 2006)
- Better mapping well-being dynamics
- Resilience targeting/ resilience mapping (Barrett and Constas 2014, Cisse & Barrett 2018, Upton, Cisse & Barrett 2016, Knippenberg et al. 2019)

Static versus dynamic innovations

- Food insecurity early warning using high frequency data (Mude et al. 2009, Lentz et al. 2018)

- Tang et al. (2018), Yeh et al. (2020) demonstrate that CNNs trained on changes in satellite imagery can predict changes in consumption or asset wealth in future periods

- Browne et al. (2021) produce contemporaneous and sequential prediction of correlated asset wealth and malnutrition indicators

Static versus dynamic frontiers

- High frequency data needed for monitoring and early warning (Barrett 2010, Headey & Barrett 2015)

- Focus on prediction of changes (not just levels)

- Integration of time series statistics with ML tools with application in these settings

Data

- Undersupply of the global public good of collection, standardization, updating and open access curation of key variables

- One can only reliably predict states and processes that have been previously observed in data → assumed stationarity in DGP

⟶ COVID has likely accelerated trends towards more creative data collection (Blumenstock 2020)

Summary

- Big data and ML methods are revolutionizing mapping, targeting, M&E, and early warning

- However, effective use requires thoughtful consideration of the purpose and use cases of the map/tool/model

- Data availability and curation remain a serious limitation

Full article available at ***Applied Economics Perspectives & Policy***

Poverty Prediction with Vegetation Index

Binh Tang [1] Yanyan Liu [2] David Matteson [1]

[1]Cornell University
[2]International Food Policy Research Institute

July 28, 2021

1.1 Motivation

- Recent advances in remote sensing and machine learning have opened up a new path for poverty prediction.
  - Jean et al. (Science 2016) uses static Google images and nighttime lights data to predict consumption expenditure and asset index at the community level.
  - Yeh et al. (Nature Communications 2020) uses dynamic Google images to predict asset index both over space and over time also at the community level.

Key limitations of using Google images

- Google images cannot capture crop and rangeland conditions.
  - Poor predictive performance among the very poor communities that rely heavily on agriculture as the main income source
- Google images change slowly over time
  - Poor performance for sequential prediction ($r^2 = 0.15$ in Yeh et al.)

Key limitations of using Google images

- Google images cannot capture crop and rangeland conditions.
  - Poor predictive performance among the very poor communities that rely heavily on agriculture as the main income source
- Google images change slowly over time
  - Poor performance for sequential prediction ($r^2 = 0.15$ in Yeh et al.)

1.2 Overview of our study

- We leverage the normalized difference vegetation index (NDVI) at 250-meter resolution to estimate consumption expenditure and asset index at the community level both over space and over time.
- NDVI provides a signal for crop heath and rangeland conditions. Different from Google images, NDVI is fast-changing and correlated with well-beings among the very poor.
- Methodologically, we follow Jean et al. (2016) to rely on convolutionary neural networks (CNNs) and transfer learning using the nightlights data as intermediate labels.

1.3 Key findings

- Prediction over space:
  - The overall predictive power of our NDVI-based model is comparable to Google images-based model
  - Our NDVI-based model outperforms Google images-based model in predictive performance among the poorer clusters.
- Prediction over time:
  - Prediction on consumption expenditure for an out-of-sample period in Uganda
  - NDVI can effectively detect consumption variation over time among the poor communities.

1.3 Key findings

- Prediction over space:
  - The overall predictive power of our NDVI-based model is comparable to Google images-based model
  - Our NDVI-based model outperforms Google images-based model in predictive performance among the poorer clusters.
- Prediction over time:
  - Prediction on consumption expenditure for an out-of-sample period in Uganda
  - NDVI can effectively detect consumption variation over time among the poor communities.

1.4 Contributions

- We demonstrate that with CNNs, publicly available, moderate-resolution NDVI images alone can predict poverty measurements as accurately as state-of-art methods using Google's high resolution images with more than 4000 features.
- Our model improves predictive accuracy among the poorest rural communities.
- To our knowledge, our study is the first attempt to predict future-period consumption expenditure.
  - Sequential prediction of consumption is especially meaningful for early warning, because consumption captures current poverty (McBride et al. 2021).

1.4 Contributions

- We demonstrate that with CNNs, publicly available, moderate-resolution NDVI images alone can predict poverty measurements as accurately as state-of-art methods using Google's high resolution images with more than 4000 features.
- Our model improves predictive accuracy among the poorest rural communities.
- To our knowledge, our study is the first attempt to predict future-period consumption expenditure.
  - Sequential prediction of consumption is especially meaningful for early warning, because consumption captures current poverty (McBride et al. 2021).

1.4 Contributions

- We demonstrate that with CNNs, publicly available, moderate-resolution NDVI images alone can predict poverty measurements as accurately as state-of-art methods using Google's high resolution images with more than 4000 features.
- Our model improves predictive accuracy among the poorest rural communities.
- To our knowledge, our study is the first attempt to predict future-period consumption expenditure.
  - Sequential prediction of consumption is especially meaningful for early warning, because consumption captures current poverty (McBride et al. 2021).

Outline

1. Introduction
2. Data
3. Method
4. Results
   - Prediction over space
   - Prediction over time
5. Conclusion

Outline

1. Introduction
2. Data
3. Method
4. Results
   - Prediction over space
   - Prediction over time
5. Conclusion

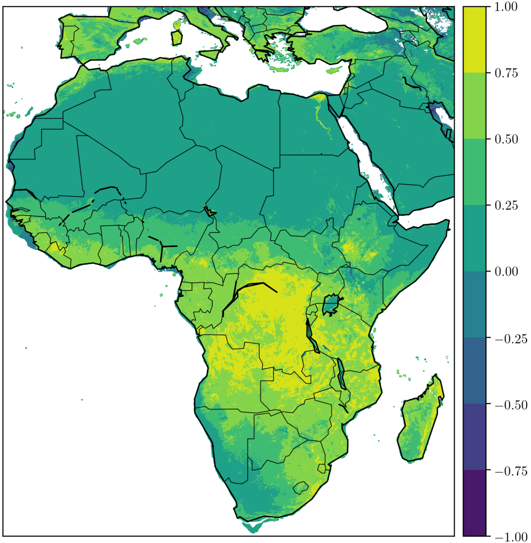
2.1 Poverty measures

- Consumption expenditure at the community/cluster level from Living Standards Measurement Studies (LSMS)
  - Current poverty measure (McBridge et al. 2021)
  - Malawi 2013, Nigeria 2013, Tanzania 2012, Uganda 2011 & 2013
- Asset index at the cluster level from Demographic and Health Surveys (DHS)
  - Chronic poverty measure (McBridge et al. 2021)
  - Malawi 2010, Nigeria 2013, Rwanda 2010, Tanzania 2010, Uganda 2011

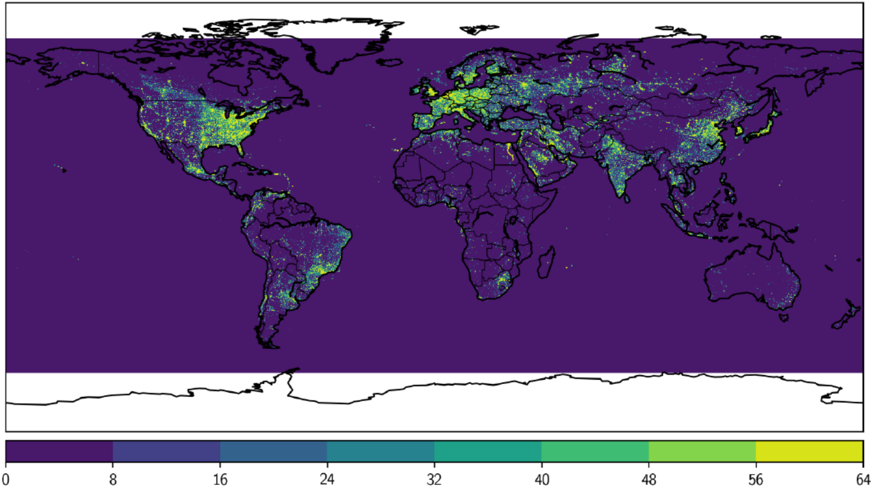
2.2 Predictors

- Annual NDVI images at the 250-m resolution
- Nightlights data - intermediate labels for transfer learning
- Google daytime images - for comparison with our NDVI-based model

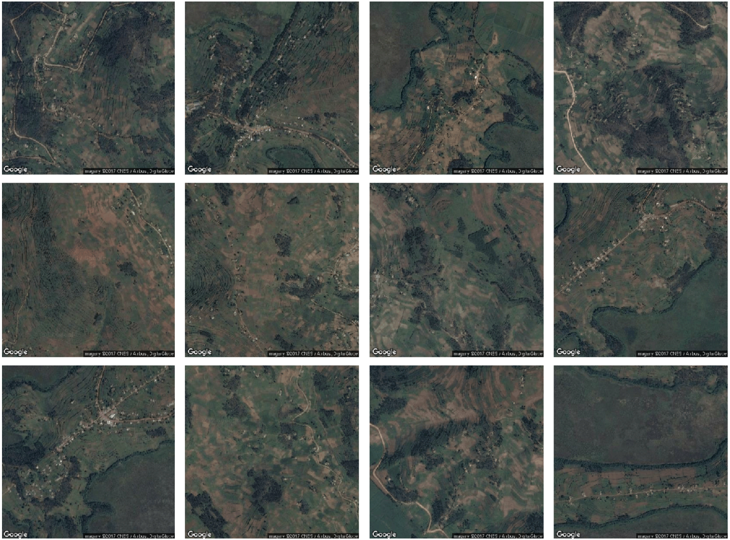
NDVI in 2010



Night lights in 2010



Sample of daytime satellite images from Google



Outline

1. Introduction
2. Data
3. Method
4. Results
    - Prediction over space
    - Prediction over time
    - Robustness check using SIF
5. Conclusion

3. Method

- We apply CNNs and transfer learning, following a two-step procedure to bypass the lack of labeled responses.
    1. Fine-tune a VGG-16 network on NDVI images to predict nighttime light intensities, in order to extract the NDVI features.
    2. Fit random forest models using these NDVI features to predict consumption and asset at the cluster level.
- The combination of NDVI images and nightlights allows vegetation features indicative of economic activity to be learned and generalized to the poverty prediction task.

Outline

1. Introduction
2. Data
3. Method
4. Results
   - Prediction over space
   - Prediction over time
5. Conclusion

4.1 Prediction over space

- Individual country analysis
- 5-fold cross validation

Table: Consumption: Spatially cross-validated R^2 of NDVI versus Google images (Jean et al.)

Country	Year	Google images	NDVI
Malawi	2013	0.37	0.341 (0.038)
Nigeria	2013	0.42	0.387 (0.013)
Tanzania	2012	0.55	0.603 (0.019)
Uganda	2011	0.41	0.490 (0.012)

Notes: Standard errors of R^2 in the parentheses.

Prediction over space - asset index

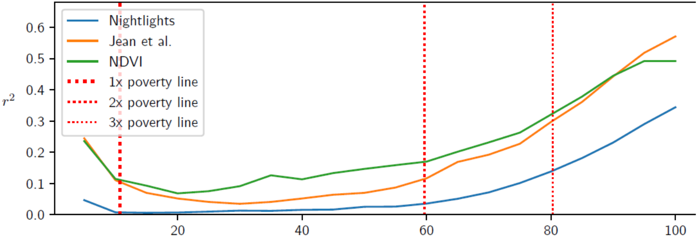
Table: Asset: Spatially cross-validated R^2 values NDVI versus Google images (Jean et al.)

Country	Year	Google images	NDVI
Malawi	2010	0.55	0.498 (0.020)
Nigeria	2013	0.68	0.738 (0.005)
Rwanda	2010	0.75	0.725 (0.022)
Tanzania	2010	0.57	0.638 (0.012)
Uganda	2011	0.69	0.751 (0.007)

Notes: Standard errors of R^2 in the parentheses.

Predictive performance by poverty level - Consumption

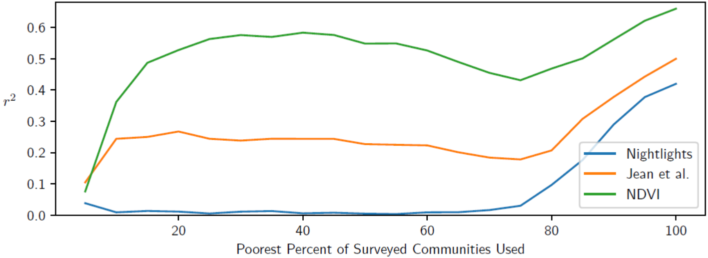
Figure: Consumption: R^2 by poverty, NDVI versus nightlights and Google images (Jean et al.)



Notes: Using pooled observations across the four LSMS countries, we run separate trials for increasing percentages of the pooled dataset (e.g., the x-axis value of 60 indicates all surveyed communities below the 60th percentile of consumption are included.)

Predictive performance by poverty level - Asset index

Figure: Asset: R^2 by poverty, NDVI versus nightlights and Google images (Jean et al.)



Notes: Using pooled observations across the five DHS countries, we run separate trials for increasing percentages of the pooled dataset (e.g., the x-axis value of 60 indicates all surveyed communities below the 60th percentile of asset are included.)

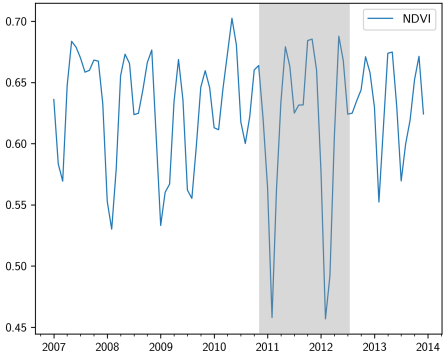
4.2 Prediction over time

- 209 communities in Uganda surveyed in 2011-2012 and 2013-2014
- Train the random forests model on the 2011 NDVI feature maps and test the model on the updated NDVI features in 2013.

4.2 Prediction over time (2)

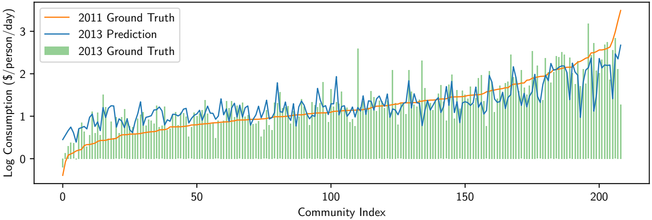
- The 2011-2012 East Africa drought affected a large area of Uganda, as shown in the figure:

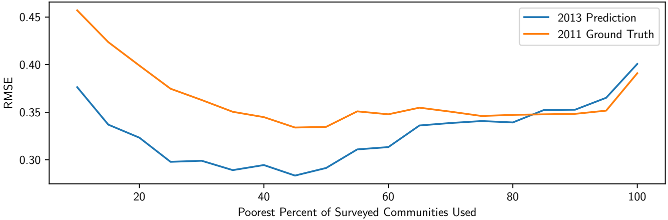
Figure: NDVI measurements for Uganda 2007-2014



4.2 Prediction over time (3)

Figure: Consumption: Sequential predictions for LSMS communities in Uganda





Outline

1. Introduction
2. Data
3. Method
4. Results
    - Prediction over space
    - Prediction over time
5. Conclusion

5.1 Summary of findings

- Prediction over space:
    - Overall predictive power of NDVI is comparable to Google images.
    - NDVI outperforms Google images in predictive power among the poorer clusters.
- Sequential prediction: NDVI can effectively detect consumption variation over time among the poor communities.

5.2 Policy takeaways

1. It is important to include predictors that can capture crop and rangeland conditions if we intend to capture poverty dynamics among the rural poor communities
   - Careful selection of predictors to serve the study purpose
2. Potential to reduce the computational dimension using alternative datasets without sacrificing predictive power
   - Lower the technical bars
3. Advanced machine learning methods (such as CNN) combined with suitable data (such as NDVI) can push forward the frontier of the research on poverty mapping, early warning, and program impact evaluation.
   - Effective collaboration between economists, computer scientists, remote sensing scientists is key!

5.2 Policy takeaways

1. It is important to include predictors that can capture crop and rangeland conditions if we intend to capture poverty dynamics among the rural poor communities
   - Careful selection of predictors to serve the study purpose
2. Potential to reduce the computational dimension using alternative datasets without sacrificing predictive power
   - Lower the technical bars
3. Advanced machine learning methods (such as CNN) combined with suitable data (such as NDVI) can push forward the frontier of the research on poverty mapping, early warning, and program impact evaluation.
   - Effective collaboration between economists, computer scientists, remote sensing scientists is key!

5.2 Policy takeaways

1. It is important to include predictors that can capture crop and rangeland conditions if we intend to capture poverty dynamics among the rural poor communities
    - Careful selection of predictors to serve the study purpose
2. Potential to reduce the computational dimension using alternative datasets without sacrificing predictive power
    - Lower the technical bars
3. Advanced machine learning methods (such as CNN) combined with suitable data (such as NDVI) can push forward the frontier of the research on poverty mapping, early warning, and program impact evaluation.
    - Effective collaboration between economists, computer scientists, remote sensing scientists is key!

Multivariate Random Forest Prediction of Poverty and Malnutrition Prevalence

July 26, 2021

Overview

- Want to predict / estimate poverty and malnutrition prevalance in developing countries for targeting, monitoring, early warning
- Historical efforts rely on expensive and imprompt surveys
- Explore use of machine learning and remote sensing (RS) to generate cheaper and higher spatial / temporal resolution estimates

Data overview

- We use open source data to ensure accessibility
- 5 DHS derived outcomes (asset poverty prevalence, child wasting, etc.)
- RS Features / inputs:
  - Meteorological and agricultural data (SIF, CHRPS, LST)
  - Conflict Data
  - Basic location / remoteness data (e.g. lat,lon, alt)
  - IFPRI market food price data

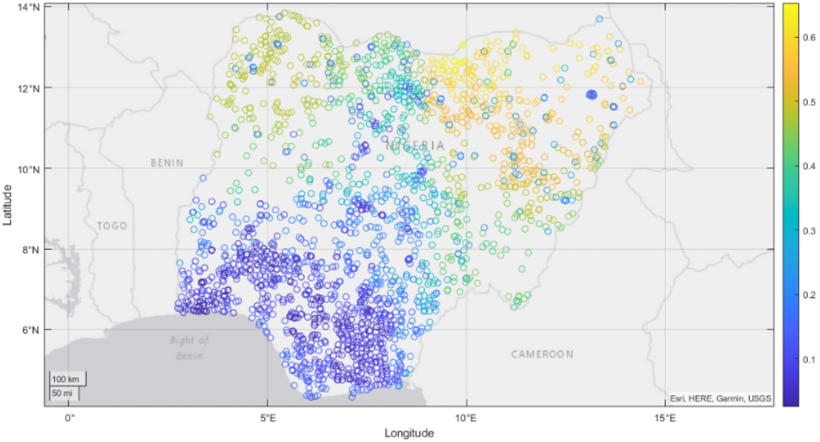
Method Overview

- State of art related works typically employ deep / transfer learning
- Test whether easier to implement / more interpretable methods could be used instead
- Since outcomes may be correlated, test whether a joint (multivariate) random forest model does better than independent modeling of outcomes
- Identify key features for prediction

Predictive Frameworks

- DHS surveys are staggered across 11 countries and over roughly 1 decade
- Food price data varies across countries, so training and testing done on country specific basis
- Two predictive tasks:
  - Contemporaneous prediction: Predict poverty and malnutrition prevalence in unsurveyed spatial locations using historical and current survey data
  - Nowcasting: Generate short term future forecasts for prevalence using historical survey data and present RS data (features)
- First is for generalizing survey derived prevalence estimates to unsurveyed regions, second is for short term forecasting

Spatial Forecasts



Predictive Framework

- Predictive accuracy is assessed at
  - Aggregate level (pooling all surveys)
  - Country level (Pooling all surveys within each country)
  - Individual survey level
- Accuracy measured by (out of sample) r^2 and RMSE normalized by observed prevalence range

Predictive Results Summary

- Contemporaneous prediction is easier than nowcasting
- Asset poverty easier to predict than malnutrition prevalence
- Joint and independent random forests do about as well as deep learning methods for prevalence prediction, when results are assessed in aggregate or at country-specific levels
- Do worse at "early" survey level due to small sample sizes (transfer learning is helpful)

Survey size vs individual survey performance

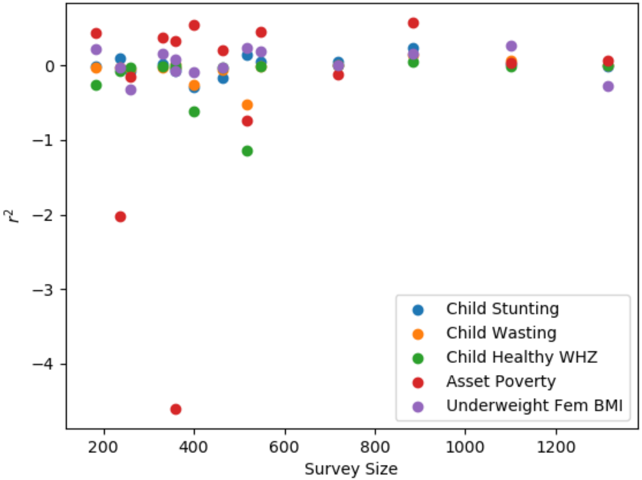


Fig. Nowcasting

Predictive Results Summary

- Exact performance measured by r^2 varies based on prediction regime, pooling, and target outcome
- NRMSE stays quite low, 10-20 percent of observed indicator range
- So relatively accurate prevalence estimates can be generated with random forests, for all predictive frameworks
- Joint prediction confers some mild benefits when nowcasting

Variable Importance

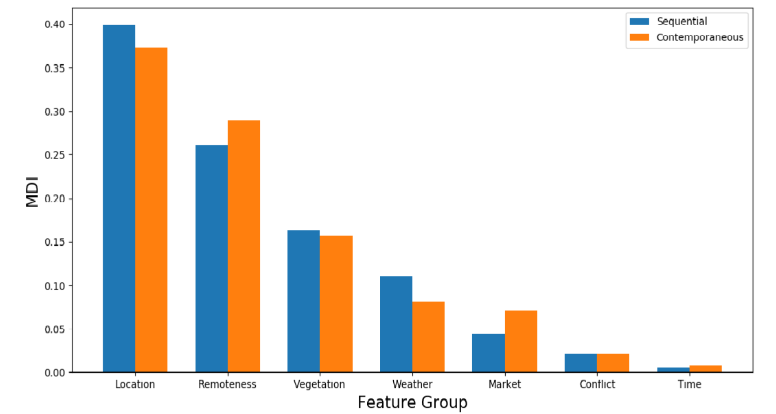


Fig. Mean MDI over all surveys, grouped by data type. Location refers to survey latitude, longitude, altitude, and slope. Remoteness indicates urban-rural status and distance to nearest major city. Vegetation includes pasture coverage, tree coverage, and SIF readings. Weather includes CHRPS and LST data. Market contains all food price data. MDI for each data type is computed by combining individual feature MDIs for each feature within that category.

Variable Importance

- Basic location data is most useful
  - Effective for short to medium term targeting
  - May work less well in heterogeneous areas or for extrapolation
- Meteorological and agricultural data useful
- Food price data unimpactful, more useful if collection becomes more standardized and more markets observed

Conclusions

- Random forests and open source, RS data can generate tolerably accurate estimates of poverty and malnutrition prevalence
- Accuracy of estimates is on par with deep learning methods for large data, outperformed on small data
  - Enhanced focus on surveys to augment statistical methods
  - Surveys can't be phased out yet
- Better performance for contemporaneous prediction of asset poverty
- Joint estimation may be helpful for nowcasting

About the Project Website

Medha Bulumulla

http://barrett.dyson.cornell.edu/research/innovations.html

Dyson
Cornell
SC Johnson College of Business

Home About Research Outreach Teaching Search

USAID Feed the Future Global Poverty and Malnutrition Estimation Project

We combine the latest advances in satellite vegetation remote sensing and physical measurement with other publicly available data, processed using accessible machine learning techniques, to explore the potential to provide ***accurate, timely, and lower cost*** monitoring of key Feed the Future (FtF) outcome indicators such as asset poverty and nutritional status at the sub-national community level in low-income FtF countries. The resulting indicators enable higher frequency monitoring and adaptive targeting, as well as careful impact evaluation of FtF and other interventions when combined with rigorous research design around project or program participation

USAID FROM THE AMERICAN PEOPLE

We thank the United States Agency for International Development for financial support under cooperative agreement # 7200AA18CA00014, “Innovations in Feed the Future Monitoring and Evaluation - Harnessing Big Data and Machine Learning to Feed the Future”. All data and written products are solely the authors’ responsibility and do not necessarily reflect the views of USAID or the United States Government.

Research Team

Research Team

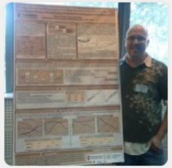

Christopher Barrett
Professor
Dyson School of Applied Economics and Management
Cornell University


Chris Browne
PhD candidate
Applied Math
Cornell University


Medha Bulumulla
Undergraduate
Statistics
Cornell University


Susana Constenla
Research Analyst
International Food Policy Research Institute

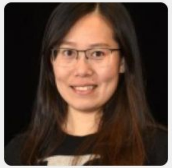

Leiqiu Hu
Assistant Professor
Atomspheric and Earth Science
University of Alabama-Huntsville


Oz Kira
Postdoctoral Associate
School of Integrative Plant Science
Cornell University


Yanyan Liu
Senior Research Fellow
International Food Policy Research Institute


David Matteson
Associate Professor
Statistics and Data Science
Cornell University


Linden McBride
Assistant Professor
Economics
St. Mary's College of Maryland


Ying Sun
Assistant Professor
School of Integrative Plant Science
Cornell University


Jiaming Wen
Graduate Student
School of Integrative Plant Science

Papers

Papers

A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF)

Jiaming Wen, Philipp Köhler, Grégory Duveiller, Nicholas C. Parazoo, Troy S. Magney, Giles J. Hooker, L.Yu, Christine Y. Chang, Ying Sun

Remote Sensing of Environment

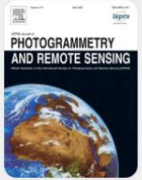
20 January 2020

High-Resolution Global Contiguous SIF of OCO-2

L. Yu, Jiaming Wen, Christine Y. Chang, Christian Frankenberg, Ying Sun

Geophysical Research Letters

11 December 2018

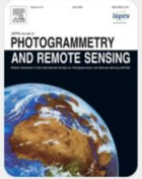


Improved estimates of monthly land surface temperature from MODIS using a diurnal temperature cycle (DTC) model

Leiqiu Hu, Ying Sun, Gavin Collins, Peng Fu

ISPRS Journal of Photogrammetry and Remote Sensing

19 August 2020



Extraction of sub-pixel C3/C4 emissions of solar-induced chlorophyll fluorescence (SIF) using artificial neural network

Oz Kira, Ying Sun

ISPRS Journal of Photogrammetry and Remote Sensing

21 January 2020

Presentations

Presentations

USAID Roundtable Presentation

24 May 2021

USAID Project Presentation

30 March 2020

Forecasting poverty and malnutrition for early warning, targeting, monitoring, and evaluation

4 January 2021

Can Publicly Available Data and Machine Learning Accurately Predict Malnutrition and Poverty?

8 May 2020

Webinar 1: Better Measures of Land Surface Temperatures and Solar-Induced Chlorophyll Fluorescence (SIF) to Improve Monitoring for Drought-Stressed Crops and Crop Productivity

Register Here

Tuesday July 20, 10:00-11:30 AM

Webinar 2: Using Big Data and Machine Learning to Predict Poverty and Malnutrition for Targeting, Mapping, Monitoring, and Early Warning

Register Here

Tuesday July 28, 8:00-9:30 AM

Project Reports and Data Sources

Project Reports

Innovations in Feed the Future Monitoring and Evaluation: Harnessing Big Data and Machine Learning to Feed the Future: Annual Performance Report

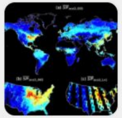
October 2020

Development Experience Clearinghouse

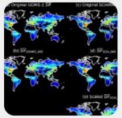
Data Sources

Monthly LST Products over the Feed-the-Future Countries

We focus on data from eleven USAID Feed the Future (FTF) priority countries: Bangladesh, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Mali, Nepal, Nigeria, Senegal, and Uganda.



High Resolution Global Contiguous SIF Estimates Derived from OCO-2 SIF and MODIS



Harmonized SIF from GOME-2 and SCIAMACHY

This dataset was created by fusing SIF retrievals from SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) and Global Ozone Monitoring Experiment 2 (GOME-2) onboard MetOp-A developed at German Research Center for Geosciences (GFZ)

Data Source Table

Code and User Guide

Code

Translational User Guide

A guide to the Python code written by Chris Browne from Multivariate random forest prediction of poverty and malnutrition prevalence and annotated code written by Medha Bulumulla.

Translational User Guide

This page provides the code and annotated code to explain the Python routines used to generate variables and to execute the predictions in Multivariate random forest prediction of poverty and malnutrition prevalence(2021). Here you can learn how to recreate the code for your needs.

The annotated guides can not compile. They contain explanations that are not commented. If you want the code to compile, please download the desired Stata, R, or Python files.

How to Use the Translational User Guide

- Download both the annotated guide(Word document) and compilable file(Python file)
- In the annotated document, the black text in Times New Roman is the annotation and the grey font is the actual code.
    - You could delete the annotations and copy it to a Python file, but there is a compilable file available.
- The annotations will explain what the code is doing and assist in recreations

Annotated Guide

Uncompilable Word Document

Python File

Complilable File

Contemporaneous Forecasting

Correlated Random Forest Sequential Nowcasting and Contemperous Prediction using Python

Code written by Chris Browne.

Contemporaneous Forecasting

Annotated Guide

Uncompilable Word Document

Python File

Complilable File

```
##set seed
np.random.seed(6023)
```

Setting the Forest Size

```
nt = 2000
```

Five Fold Cross Validation Testing and Training are performed 5 times for each survey

```
folds = 5
```

Creating a vector that has different random forest methods: independent(ind) or joint

```
rftypes = ['ind','joint'] ##ind must be first
```

Setting the countries, these might be different for you.

```
countries = ["Bangladesh","Ethiopia","Ghana","Guatemala","Honduras","Mali","Nepal","Kenya","Senegal","Uganda","Nigeria"]
```

Setting the years, these might be different for you.

```
pyears = [["04","07","11","14"],["05","11","16"],["08","14"],["14"],["11"],["06","12"],["06","11","16"],['08','14'],["05",'10'],["06",'11','16'],
["08","13"]]
```

Sequential Nowcasting

Sequential Nowcasting

Annotated Guide

Uncompilable Word Document

Python File

Complilable File

Sequential Now-casting: To be used in Early Warning Systems.

```
import numpy as np
import pandas as pd
import pickle
import pdb
from sklearn.ensemble import RandomForestRegressor as RF
from sklearn.metrics import r2_score as r2
from sklearn.metrics import mean_squared_error as mse
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Ridge as ridge
```

Setting the seed ensures that the results will be the same when reproduced.

```
##set seed
np.random.seed(9823)
```

Setting the Forest Size to 2000

```
nt = 2000
```

Creating a vector that has different random forest methods: independent(ind) or joint

```
rftypes = ['ind','joint'] ##ind must be first
```