FEED THE FUTURE

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

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

Annual Performance Report

October 2020

Reporting Period: August 17, 2018 (inception) – September 30, 2020

AOR/COR/Activity Manager Name & Office: Dr. Farzana Ramzan, Bureau for Resilience and Food Security, Office of Strategic Planning and Performance Management

Implemented by: Cornell University

Partners: International Food Policy Research Institute (IFPRI), University of Alabama-Huntsville (UAH)

PI: Prof. Christopher B. Barrett

Co-PIs: Prof. Leiqiu Hu (UAH), Dr. Yanyan Liu (IFPRI), Prof. David Matteson (Cornell), Prof. Linden McBride (St. Mary's College), Prof. Ying Sun (Cornell)

Innovations in Feed the Future Monitoring and Evaluation Cooperative Agreement #7200AA18CA00014 between the United States Agency for International Development (USAID) and Cornell University

Harnessing Big Data and Machine Learning to Feed the Future

Cornell University

PI: Christopher B. Barrett; Co-PIs: Leiqiu Hu (UAH), Yanyan Liu (IFPRI), David Matteson, Ying Sun

Annual Progress Report, Project Inception (August 17, 2018) – September 30, 2020

We are now roughly two-thirds of the way through the three-year award period. We completed both administrative and technical activities necessary to accomplish the project's stated objectives to date.

Administrative actions completed include:

- Executed award contract between Cornell University and USAID
- Held post-award kick-off meeting (virtually) with USAID (September 20, 2018)
- Executed subawards with the University of Alabama – Huntsville (UAH) and the International Food Policy Research Institute (IFPRI)
- Executed individual consulting contracts with Dr. Linden McBride (St. Mary's College)
- Completed and filed the required Activity Monitoring, Evaluation, and Learning (MEL) Plan

Technical activities completed met the milestones indicated in our original proposal and MEL. These can be broken into four distinct areas of work, as described below. We have generally adhered to our workplan, although our modeling work has adapted as we learned from our initial work. We have faced some delays on specific indicators and countries due to difficulties accessing some data series (described below), or in accessing the geolocations associated with series, as is necessary for the spatially explicit predictive work at the heart of our project. The COVID-19 pandemic caused some significant delays due to unexpected demands on team members' time to attend to family care issues. The pandemic has also necessitated cancellation of planned travel and in-person training sessions. Overall, the project has progressed reasonably well.

1. **Development of SIF data product:** We have developed and published a novel data set on solar-induced chlorophyll fluorescence (SIF), an innovative metric of photosynthesis that can be measured from satellite sensors. SIF emissions carry functional information about the metabolic and physiological states of plants. These novel measures offer new opportunities to measure crop productivity. A major prior gap in the SIF data products available had been low spatial resolution and sparse data acquisition. Our project has worked to remedy those shortfalls under the direction of co-PI Sun.

   Per Task 1a ("Develop and validate a SIF-based modeling framework to estimate crop yield") our group developed two high-resolution satellite SIF products at 5km resolution using machine learning techniques. The first product used a hybrid approach that combines artificial neural network (ANN) algorithms and physiological constraints to gap-fill the native SIF measurements from the Orbiting Carbon Observatory (OCO)-2 satellite mission for the period 2014-2018 at bi-weekly resolution. The physiological constraint is implemented by stratifying time and biome type, considering that the relationship between predictors and response variables are physiologically different with time and biome types. This work, funded on a separate grant, yielded a paper published in *Geophysical Research Letters* (Yu et al. 2018). This product is highly consistent with independent airborne measurements at high resolution. It is capable of successfully identifying highly productive agricultural sectors in a more spatially explicit way than the original SIF, and can advance drought monitoring and mitigation in a physiologically meaningful manner.

   The work that was already underway led directly to the second product, developed under this project: a long-term (2002-present) harmonized SIF dataset at monthly resolution created by fusing multiple satellite SIF utilizing ANN and Random Forest (RF), respectively, and cumulative probability distribution (CDF) matching (Wen et al. in review). The key advantage of this product is that it is the first long-term SIF product, overcoming the relative short lifetime of individual satellite missions that have SIF capability. This product allows examination of the historical variation of vegetation activities using SIF that was not previously possible. Machine learning techniques offer a flexible and computationally efficient way to generate high-resolution SIF. In addition, we have quantified the data uncertainty of this SIF product to facilitate an eventual estimation of the confidence level of the predicted FtF indicators, or other outcome variables predicted using these SIF series. Furthermore, we have investigated the capability of this developed SIF product in characterizing drought events and found that it can reveal physiological response to drought, which is beyond what the conventional vegetation indices can offer. This paper has been published in *Remote Sensing of Environment* (Wen et al., 2020), the top remote sensing journal.

Realizing that the 5km spatial resolution of satellite SIF observations usually consist of mixed SIF signals contributed by different crop types with distinct phenology (modulated by management practices) and varying SIF emission capacities, we developed a sub-pixel SIF extraction framework for corn and soybean using the US Corn Belt as a case study. Our results demonstrated that the estimated sub-pixel SIF could successfully reproduce the original pure SIF values constituting the mixed pixel, with a normalized root mean squared error (NRMSE) of <10% during the peak growing season. We further demonstrated that this framework substantially outperforms the parsimonious linear extraction methods. This developed sub-pixel SIF extraction framework was then applied to generate regional-scale SIF maps for corn and soybean at 0.05° in the US Midwest. Although tested for corn and soybean only in the US, the developed framework has the potential to resolve sub-pixel SIF of more endmembers from coarse SIF observations elsewhere in the world. This paper has been published in *ISPRS Journal of Photogrammetry and Remote Sensing* (Kira and Sun, 2020).

Further, we have used these SIF products to estimate crop yields using the US Midwest as a testbed, where rich crop yield datasets are available from USDA NASS and the landscape is relatively more homogeneous than in FtF countries. We have tested a number of machine learning models from 16-day SIF datasets against the end-of-year crop yield. We found encouraging results: 1) SIF can outperform conventional vegetation indices (e.g., NDVI, EVI) in yield prediction with improved R^2 and reduced RMSE; 2) if separating crop types, the SIF-based yield prediction can be further improved especially for corn. The robustness of the model was tested by out-of-sample prediction. Currently, we are incorporating a new satellite SIF product, TROPOspheric Monitoring Instrument (TROPOMI) onboard the Copernicus Sentinel-5 Precursor satellite, to our prediction framework, because it has higher temporal resolution (lower spatial resolution and shorter temporal coverage than our developed two SIF products). Moreover, we attempt to develop an unsupervised yield prediction approach based on the mechanistic modeling of SIF and photosynthesis developed by Co-PI Sun's group (Han et al., 2020, under review). With these new efforts, we plan to submit this working paper to *Environmental Research Letters* by early 2021. Currently, we are also adopting this unsupervised SIF-based approach to Ethiopia, one of the FtF countries, to test its scalability and robustness for estimating yields of corn and wheat.

In May 2020, Co-PI Ying Sun also published an AgriLinks blog post, "Measuring Plant Photosynthesis from Space to Understand Crop Production on the Ground: The Promise of Solar-Induced Chlorophyll Fluorescence" to report on these new advances in SIF data products and their performance in predicting crop yields.

The two high-resolution SIF products have been made publicly available at Cornell Box. Users can access these datasets through Co-PI Ying Sun's Cornell website as a gateway https://www.yingsun.info/datasets-and-tools. In addition, the first product by Yu et al. (2019) has also been archived at Oak Ridge National Lab (ORNL) DACC https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1696

2. **Development of LST data product:** Land surface temperature (LST) is an effective indicator of land-atmosphere interactions. The UAH-based team led by co-PI Hu has generated a new, quality-controlled LST data series over the FtF countries to identify unfavorable crop growing conditions, such as droughts. In order to fulfill the goal to produce high-quality and temporally representative data with a long record (2003-present) at a high spatial resolution of 1km, we combined two major satellite-based LST products: geostationary observations (Copernicus global LST products that are consolidated from multi-geostationary satellites at 5 km with an hourly frequency, 2010-Present) and Sun-Synchronous-orbiting satellites (MODIS LST products at 1 km with four observations per day, 2003-Present). The geostationary LST products feature a high temporal resolution at hourly frequency but rather coarse spatial resolution at 5km and a shorter archived period from 2010-present. First, we fit the diurnal temperature cycle model (DTC) using the high-frequency geostationary LST and then apply the geostationary-based DTC parameters to reconstruct the diurnal cycles of the less frequent sun-synchronous satellite observations at a monthly scale. The process results in the bias-corrected monthly mean and maximum LST at 1km from 2003-2018 over the FtF countries. The advantage of the combination of two different types of LST products is to reduce and constrain the unknown parameters in the DTC model and to help yield a better-quality controlled data with a longer period that matches SIF products (2003- present) and a finer spatial resolution at 1km.

   Toward this end, we developed a data processing framework to implement the physics-based DTC model with an optimized scheme using infrequent MODIS observations at 1km (2003-2018). The DTC model effectively produces monthly diurnal temperature from MODIS and geostationary observations and the DTC-model results are more accurate than the other existing approaches or satellite products which do not employ temporal gap filling.

   The quality of DTC-based LST products were assessed in independent sources of observations. First, we have validated the accuracy of DTC-based monthly LST by observations from in-situ ground flux sites globally from FLUXNET, a global network of micrometeorological tower sites. Due to the absence of FLUXNET sites over the target FtF countries, we are only able to compare the LST data with the locations that have ground observations to assess the applicability of the DTC approach over different climate regions and various land cover types. This, however, serves to validate the quality-controlled diurnal data series we have constructed over a long period at reasonably high spatial resolution. We found a strong agreement between ground observations and DTC-based monthly mean LST among various land cover types with high correlation coefficients consistently greater than 0.95. The high temporal consistency gives us confidence about our ability to use this method to identify temperature anomalies in growing seasons using this new data product. The DTC-based monthly LST shows a nearly-zero mean difference (0.3 °C) against the ground observations and has an overall RMSE of about 2.2°C. Moreover, the data quality has a seasonal variation. During the growing (warm) seasons, we observed an overall better agreement between satellite estimation and ground observations. However, the RMSE varies depending on the locations.

There is an inherent incomparability between satellite observations and ground-based measurements. First, system biases exist among different satellite sensors (e.g., MODIS, GEOS, and MSG) and the ground-based pyrgeometer. For example, the satellite LST retrievals rely on multiple narrow thermal bands in each sensor and involve various approximations, e.g., estimating the broadband radiation for LST from the measured narrowband irradiance, simplification and assumptions for surface emissivity and parameters for atmospheric attenuation correction. Differently, pyrgeometer directly measures the broadband upwelling radiation at the ground level, and the atmospheric impact is trivial due to the very close distance from the target.

Second, there are mismatched spatial footprints between satellite grid and in-situ measurements. The ideal comparison requires the site has a greater homogeneous surface, which is less likely true over the flux sites due to the complex nature of surfaces.

Third, we compared the all-sky LST from ground observations, but satellite products represent pseudo-clear sky conditions, which does not include the rain days. These partly contribute to the uncertainties in the validation.

In order to overcome these aforementioned limitations and to better illustrate the quality of our LST products over the FtF countries with good spatial representation, we further compare our approach to the simple composite approach for MODIS suggested by Chen et al. (2017) and mean composite directly from hourly observations from geostationary satellites at 5km. Our proposed DTC method outperforms the other two against the ground observations.

Furthermore, besides the monthly mean LST, we are able to estimate the representative maximum temperature from infrequent MODIS observations through the diurnal reconstruction. In order to identify the better predictable parameter for drought conditions, we linked the temporal thermal anomalies of monthly mean and maximum LST with the soil moisture anomalies simulated from three NLDAS Phase 2 (NLDAS-2) models over the contiguous United States (CONUS). The comparison unveils the monthly maximum temperature responds to droughts more sensitive than the mean value. Thus, the maximum approximation of monthly temperature over the FtF countries has been used in the ML model.

We have published the manuscript that summarizing the new data products at the *ISPRS Journal of Photogrammetry and Remote Sensing in 2020*. The relevant results were also presented at the 2020 American Meteorology Society annual meeting in

Boston. A blog describing the new LST series, their validation, and the new opportunities they afford spatial analysts was shared with the USAID has been published on agrilinks.

The long-term monthly mean and maximum LST products over the Feed-the-Future countries will be publicly available by the summer of 2021. The products include the DTC-based LST from MODIS at 1km from 2003-2018 and DTC-based LST from geostationary observations at 5 km from 2011-2018.

3. **Collection of poverty, nutrition, and food price data:** Co-PI Liu (IFPRI) leads this project activity. We have completed collecting the food price data available from open sources. Monthly food price data are available from FAO in all FtF countries. IFPRI's Food Security Portal (FSP) provides additional daily price data in Guatemala and Honduras. Table 1 summarizes the open-access food price data for all the FtF countries. We have also explored other data sources including RATIN and Government of Ghana. However, neither data series is open source (i.e., publicly available at no or very low cost), thus they do not meet our criteria for feature inclusion in this project.

Table 1: Summary of open access food price data

Country	# Food types	# markets	Frequency	First period	Source
Bangladesh	5	1	monthly	7/1998	FAO
Ethiopia	5	8	monthly	1/2000	FAO
Ghana	8	6	monthly	1/2006	FAO
Guatemala	7	2	monthly	1/2000	FAO
	6	1	daily	1/1/2013	FSP
Honduras	3	3	monthly	1/2000	FAO
	8	3	daily	4/30/2015	FSP
Kenya	2	5	monthly	1/2006	FAO
Mali	4	7	monthly	1/2005	FAO
Nepal	2	1	monthly	1/2005	FAO
Niger	6	6	monthly	1/1990	FAO
Nigeria	6	6	monthly	8/2003	FAO
Senegal	3	11	monthly	1/2007	FAO
Uganda	5	3	monthly	1/2006	FAO

The FtF indicators, which are the dependent variables in our predictive analyses, have been constructed based on total consumption expenditures per capita (Indicators 1 & 2), height-for-age z score (HAZ) of children under five (Indicators 3 & 4), asset index (Indicators 7 & 8), weight-for-height z score (WHZ) of children under five (Indicators 8 and 10), and body mass index (BMI) of women of reproductive age (Indicator 9). Note that we were unable to find publicly available geo-tagged data on food insecurity experience scale (FIES, Indicators 5 & 6), despite multiple queries with the data stewards at FAO and elsewhere. It appears the FIES data only began getting geo-tagged in 2018. As a result, our USAID AOR/Activity Manager approved dropping that indicator.

HAZ, WHZ, BMI, and asset index series indicators were constructed from the USAID-supported Demographic and Health Surveys (DHS) data sets. That activity was led by co-PI Prof. McBride. Women's BMI underweight data was extracted directly from the DHS. Asset index data and the child nutrition indicators, HAZ and WHZ, were extracted from a DHS data compilation provided by the Advancing Research on Nutrition and Agriculture project Phase I (ARENA-I) at the International Food Policy Research Institute (IFPRI). ARENA-I is a 3-year multiple country project in South Asia and sub-Saharan Africa, funded by Bill and Melinda Gates Foundation and CGIAR Research Program on Agriculture for Nutrition and Health (A4NH), being implemented from 2015 through 2017. One of ARENA's innovations is to merge a wide range of GIS indicators on agriculture, climate, demography and infrastructure with DHS surveys, especially for sub-Saharan Africa where latitude and longitude are recorded for most survey clusters.

We do not have the programs used by the ARENA team for data cleaning and variable construction. Therefore, to guarantee the quality of key outcome variables, we cross-checked several variables (HAZ, WAZ, BMI,) constructed from the DHS raw data and the ARENA-generated prevalence estimates against the prevalence rates published by DHS Country Final Reports.

Using the ARENA data compilation, the child nutrition indicators were estimated as follows. The cluster level prevalence of stunting was estimated as the cluster level share of children less than five years of age with height-for-age Z scores of less than 2 standard deviations below the WHO Child Growth Standard mean. The cluster level prevalence of wasting was estimated as the cluster level share of children less than five years of age with weight-for-height Z scores of less than 2 standard deviations below the WHO Child Growth Standard mean. The cluster level prevalence of healthy weight children was estimated as the cluster level share of children less than five years of age with weight-for-height Z scores of greater than -2 standard deviations below and less than 2 standard deviations above the WHO Child Growth Standard mean. Household level weights were used in estimating each of the child nutrition indicators.

The data are summarized in Table 2. Note that, although it is an FTF county, Niger is not included in the nutrition and poverty indicators extracted from DHS as GIS data are not publicly available for the Niger DHS surveys.

Due to missingness in the ARENA complication, data on the women's underweight indicator was extracted directly from the DHS. Cluster level prevalence of underweight women was estimated as the cluster level share of women ages 15 to 49 with body mass index (BMI) below 18.5 out of the total number of women ages 15 to 49. Individual weights were used in estimating the women's underweight indicator. Women's underweight data were not available for Senegal in 2012 or 2014.

In addition to the nutrition indicators detailed above, we extract asset poverty data from the ARENA DHS compilation. The asset poverty indicator is estimated as the cluster level share of households falling below the USAID FTF comparative threshold for the poorest quintile of the asset-based comparative wealth index, where the comparative threshold is defined as asset index score less than or equal to -0.9080 (FTF 2018). Household weights were used to estimate the cluster level asset poverty indicator. We note again that geo-referenced DHS data were not available for Niger; therefore, Niger is not represented in the DHS nutrition and poverty indicators.

We extracted consumption expenditure-based poverty indicators from the Living Standards and Measurement Study (LSMS) data or datasets similar to LSMS as the cluster level share of survey-weighted households falling below the $1.90 per day 2011 purchasing power parity. Use of LSMS data was limited by the availability of geotags. The available data are summarized in Table 3.

Table 2. DHS nutrition and poverty indicators; DHS source year indicated

Indicator Title	Bangladesh	Ethiopia	Ghana	Guatemala	Honduras	Kenya	Mali	Nepal	Nigeria	Senegal	Uganda
Prevalence of stunted (HAZ < -2) children under five years of age	2004, 2007, 2011, 2014	2005, 2011	2003, 2008, 2014	2014	2005, 2011	2003, 2008, 2014	2006, 2012	2006 , 2011 , 2016	2003, 2008, 2013	2005, 2010, 2012, 2014	2006, 2011, 2016
Prevalence of wasted (WHZ<-2) children under five years of age	2004, 2007, 2011, 2014	2005, 2011	2003, 2008, 2014	2014	2005, 2011	2003, 2008, 2014	2006, 2012	2006, 2011, 2016	2003, 2008, 2013	2005, 2010, 2012, 2014	2006, 2011, 2016
Prevalence of healthy weight (-2 ≤ WHZ ≤ 2) among children under five years of age	2004, 2007, 2011, 2014	2005, 2011	2003, 2008, 2014	2014	2005, 2011	2003, 2008, 2014	2006, 2012	2006, 2011, 2016	2003, 2008, 2013	2005, 2010, 2012, 2014	2006, 2011, 2016
Prevalence of underweight (BMI < 18.5) women of reproductive age	2004, 2007, 2011, 2014	2005, 2011	2003, 2008, 2014	2014	2005, 2011	2003, 2008, 2014	2006, 2012	2006, 2011, 2016	2003, 2008, 2013	2005, 2010	2006, 2011, 2016
Percentage of Households below the Comparative Threshold for the Poorest Quintile of Asset-Based Comparative Wealth Index	2004, 2007, 2011, 2014	2005, 2011	2003, 2008, 2014	2014	2005, 2011	2003, 2008, 2014	2006, 2012	2006, 2011, 2016	2003, 2008, 2013	2005, 2010, 2012, 2014	2006, 2011, 2016

Table 3 Consumption data availability

Country	Consumption data available
Bangladesh	IFPRI's BIHS* 2011/12, 2015
Ethiopia	LSMS 2011/12, 2013/14, 2015/16
Mali	LSMS 2014
Nepal	LSMS 1995/96, 2003/04, 2010/11
Niger	LSMS 2011, 2014
Nigeria	LSMS 2010/11, 2012/13, 2015/16
Uganda	LSMS 2009/10, 2010/11, 2011/12, 2013/14

Note: * BIHS: Bangladesh Integrated Household Survey.

Unlike DHS data, the consumption expenditures per capita series are not always georeferenced and, in some countries where the data have been georeferenced, the geolocations of households and enumeration areas are not publicly available. This has posed challenges in countries such as Ghana, Guatemala, Honduras, Kenya, and Senegal where we have tried multiple channels to access nationally representative household survey data sets with (even randomized offset) geolocations for enumeration area (ea) aggregates (not even specific households). We are able to construct the GPS-coordinates at the enumeration area level for Nepal by matching the names of locations with the boundary documentation we obtained from the government of Nepal, with considerably eyeball checking to correct the spelling errors of the location names. However, our efforts have proved unsuccessful in other countries. Hence we only have the limited subset of FTF countries covered in Table 3 with GPS coordinates. This makes the expenditure-based poverty measures less complete across space and time relative to the other FTF indicators we predict. We have therefore focused our predictive efforts (described below) on the DHS-derived indicators only.

The FIES data are collected by the Gallup World Poll and curated by FAO. FAO reports (private communication from Dr. Carlo Cafiero, FAO Food Security and Nutrition Statistics Team Leader), however, that FIES data have only just begun (in 2018) to be georeferenced, that Gallup does not make those geolocations available, and that there are only approximately 1,000 household observations per country. Without georeferencing or denser data, the methods we use will be infeasible for predicting FIES at subnational scale. We explored accessing other, larger, georeferenced data sets that include FIES (or FIES-like) questions so as to predict the moderate and severe food insecurity indicators. This proved unsuccessful. USAID agreed we can drop this FtF indicator from the ensemble we try to predict.

Given the absence of geo-tagged household survey expenditure data from Kenya, we also investigated the possible use of an extensive data set made available to us by the National Drought Management Authority (NDMA) of Kenya. The NDMA data include 4.5 million measures of mid-upper arm circumference (MUAC) for children across 23 districts in Kenya from 1992-2016. MUAC is not a FtF indicator but is closely related to other measures of acute malnutrition, such as WHZ.

NDMA data – like early waves of the same data series collected under the government/World Bank Arid Lands Resources Management Program – are of uneven quality and thus we first invested time to assess the quality of the NDMA data. Specifically, we compared the NDMA MUAC measure with measures of malnutrition available from the highly-regarded, widely-used DHS data over the same time period. We also compared NDMA data with records of food crises in the region.

To compare the NDMA data with the DHS data, we matched the Kenya geo-referenced clusters to a Kenya district map and then extracted the district boundaries for each DHS cluster. Using the district coded DHS data, we calculated annual, district level prevalence of wasting (WHZ<- 2) and severe wasting (WHZ< -3) using household survey weights. We then matched the DHS data with the annual, district level prevalence of malnutrition (MUAC<125mm) and severe malnutrition (MUAC<115mm) in the NDMA data. We matched the data at the District level, the lowest common geographical level for the DHS and NDMA data. The shared time periods are 2008, 2009, 2014. We produced equivalence of proportions tests for each year-district measure of malnutrition and severe malnutrition. We also performed non-parametric tests of rank correlation. Generally, we do not find a strong correlation between the measures of malnutrition in the NDMA and DHS data. In a comparison of the NDMA measures of malnutrition against records of food crises in the region over time, we do find that malnutrition and severe malnutrition rise with documented food crises. But we concluded that the weak correspondence between the NDMA and DHS data made the former, larger data set insufficiently reliable to use for the purposes of predicting FTF indicators.

4. **Machine learning estimation:** Related work (Tang et al. 2018) begun under a different award has continued and was presented at a conference, the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018). The same advanced prediction methods (transfer learning, with neural networks and random forest in particular) were investigated in this project, but for the FtF indicators and locations, and with an expanded set of predictors. Building on material presented in earlier, preliminary form in our award proposal, Tang et al. (2018) demonstrate that publicly available, moderate-resolution vegetation index data – specifically, the normalized differenced vegetation index (NDVI) – can be used with convolutional neural networks (CNN) to produce poverty estimates in low-income agrarian regions with the same or greater predictive skill than do previous efforts that do not exploit features that reflect agricultural productivity. As applied to asset indices from the DHS data and per capita expenditure data from the Living Standards Measurement Surveys (LSMS) from Malawi, Nigeria,

Rwanda, Tanzania, and Uganda, Tang et al. (2018) predict poverty as well as high-resolution images constrained by Google's licensing terms. Specifically, spatially cross-validated r^2 values of the NDVI models were significantly higher than models that used Google's daytime satellite images, including improvements above 100% for asset index prediction for regions below the 2x poverty line.

Our preliminary trials started with these previous models, variables and locations. Those initial explorations suggested that we can further improve predictive accuracy by also incorporating improved SIF and LST measures as additional predictors, by the same cross-validated r^2 and RMSE metrics. A further advantage of this method is that a continuous stream of these predictor data are publicly available, and we can update FtF indicator estimates in near-real-time in response to weather shocks by conditioning on the latest fitted model, updating newly available input predictors and thereby updating all associated model predictions. This continuous updating opens up the possibility of making dynamic poverty mapping feasible at minimal cost, but with moderately high spatial resolution and predictive accuracy to inform timely decisions by policymakers.

Building on the success of Jean et al. (2016) and Tang et al. (2018), we constructed a machine learning pipeline that combines methods from deep transfer learning for image based processing and feature extraction alongside a more traditional random forest predictive model. We found the latter method to produce performance comparable to related works using convolutional neural networks, while being far simpler to implement and update, as discussed below. Moreover, we found negligible change in performance when replacing deep learning extracted features with raw features. We therefore abandoned efforts to use deep transfer learning and have concentrated exclusively on the use of random forests. We continued to make minor modifications and adapt our methods to ongoing data preprocessing. The data series discussed in Sections 1-3, image based, fine-resolution SIF,LST, rainfall, NDVI, and nightlights data, alongside geospatial variates and market data, have been incorporated as predictors into this model.

Per task 2b, integration of food price data with remotely sensed agricultural, rainfall, and land surface temperature data was completed, along with processing and inclusion of basic geospatial (ARENA) and conflict data. Per tasks 2c and 2d, assessment of predictive ability in a machine learning framework was performed for all available FtF target countries. We undertook significant model tuning and diagnostics.

After extensive evaluations, we found that convolutional (deep learning) approaches offered minimal gains in predictive performance on our extended set of FtF indicators relative to considerably simpler predictive methods based on random forests. We therefore developed a revised pipeline combining a multi-output random forest together with a multi-task Gaussian process (GP), which jointly predicts target FtF indicators. This method yielded predictive performance comparable to – and for some indicators, better than – that of the best deep learning approaches for asset prediction, while being computationally fast and overall simpler to implement and update, while also naturally encoding predictive uncertainty per task 2d. On prediction tasks other than assets, we found considerably improvement utilizing this two stage prediction procedure.

We coded these models in Python. As a check on the reproducibility of our results and in order to make our methods and code more broadly available, we spent several months replicating the entire apparatus in R. This uncovered minor issues in the prior work that we then corrected. One significant issue the R replication raised had to do with the method for computing out-of-sample predictive performance measures, as reflected in r^2 and root mean squared error (RMSE) statistics. What had originally appeared as superior performance when we added the GP to the multi-output random forests procedure turned out to be an artifice of the computational method initially chosen. When that result proved not robust, we dropped the GP so as to simplify the process and thereby make it easier for third parties to implement. Our procedure now focuses on univariate and joint random forests methods.

PI Barrett presented preliminary results of our modeling work at a heavily-attended webinar on "Near-Real-Time Monitoring of Food Crisis Risk Factors: State of Knowledge and Future Prospects," hosted by IFPRI on May 8. That presentation, entitled "Can Publicly Available Data and Machine Learning Accurately Predict Malnutrition and Poverty?" reported our initial findings, including what we believed at the time to be considerably improved predictive performance from incorporating GP in ensemble prediction so as to harness the error correlation structure among individual indicator predictions. Our replication exercises over the summer and further diagnostics, unfortunately, uncovered that those results were spurious.

Redoing our analyses after uncovering the problem with the GP cost us a couple of months. We have now completely redone our procedures in Python and successfully reproduced them in R, although we are less confident in the canned R routines we used than in our original Python code, which we are now re-replicating with fresh coding. Our results now demonstrate that simple, joint random forest models trained on our data can produce results of comparable skill for the same DHS-based indicators as those generated by current frontier approaches that use more complex deep learning methods. The relative ease of implementing our approach is an advantage for prospective third party users, as in national statistical agencies, FEWSNet, or others. We assess our model in both a contemporaneous and sequential (forecasting FtF indicators 3-5 years ahead, using only historical data) framework. We find that while our results for contemporaneous prediction are similar to recent results found in the frontier literature, our results in the sequential setting offer a significant improvement over (relatively few) previous works. Moreover, we find joint estimation of outcomes to significantly improve our forecasts in the sequential framework over prediction of single FtF indicators. Our strong results in the sequential framework signify that modestly accurate predictions of poverty and malnutrition status can be generated years in advance, giving policy influencers early assessment of prevalence rates upon which they can act. We have, however, found that our method performs best when assessed at coarse, e.g., global or country level, scales, with our method exhibiting a large reduction in performance at the single country survey level, due to small sample size, which is less of an issue for transfer learning based methods than it is for our random forest based method. We explain our method and illustrate these findings in the core manuscript draft, which we are just now completing and anticipate submitting to a peer reviewed journal in the first quarter of the third project year.

We are also developing a paper for presentation at a January 2021 conference. Tentatively titled "Forecasting Correlated Poverty and Malnutrition Indicators for Targeting, Monitoring and Evaluation Purposes," the current draft abstract for this paper reads as follows: Recent extreme weather events, the COVID-19 pandemic and east African locust infestation of 2020, outbreaks of violence in various places, and food price shocks have all vividly demonstrated that food emergencies can arise quickly, demanding concerted policy response. High quality, subnational maps with tolerably accurate estimates of recent and current poverty and malnutrition conditions can be incredibly valuable to humanitarian and development agencies designing, targeting, monitoring and evaluating interventions in such settings. However, fielding surveys to generate the necessary data takes considerable time and money. So researchers and policymakers have been looking to earth observation and other publicly available, near-real-time data streams for information that might prove useful in providing such estimates. Recent research has established that machine learning methods applied to spatially precise remotely sensed and/or call data can predict some poverty measures with reasonably good out of sample forecasting accuracy (Blumenstock et al. *Science* 2015; Jean et al. *Science* 2016; Pokhriyal and Jacques *PNAS* 2017). To date, however, those methods and models have generally not performed well in predicting indicators of malnutrition. Simultaneously, poverty prediction has been moving away from theory-informed asset dynamics and related models towards atheoretic methods that simply offer the lowest prediction error. It is important that those who rely on such models to inform the distribution of benefits understand the tradeoffs and limitations of each approach. We report on a new machine learning based forecasting method that uses a suite of publicly available data series to estimate ensembles of poverty and malnutrition data to generate national poverty and malnutrition maps. We demonstrate, using nationally representative panel data from 12 low-income countries, that joint prediction of future poverty and malnutrition prevalence at a fairly local spatial scale (survey enumeration area) generates reasonably good predictive accuracy, and improves predictive accuracy with respect to malnutrition indicators relative to single indicator sequential forecasting. We discuss how data curation and estimation might be structured and implemented as part of a faster, cheaper approach to targeting, monitoring and evaluating interventions to address food emergencies. Finally, we assess and discuss the tradeoffs between feature sets and models informed by theory versus those selected by the minimization of prediction error.

5. **Summary:** We have developed the new data products planned under tasks 1a and 1b and have made those available to the earth observation community. We have completed most tasks on time per our workplan and are well along in completing the

remaining tasks. We expect to share an initial draft for comments with experts and USAID in the first quarter of year three. We have been disseminating research results, per Task 3, through various conference, journals, blog posts and webinars. We are ready to commence outreach and capacity building activities envisioned under Task 4 – adapted for the pandemic environment – and to pursue publication of the ultimate prediction methods and results papers from this project in the new project year.

6. **References**

Chen, X., Su, Z., Ma, Y., Cleverly, J., Liddell, M., 2017. An Accurate Estimate of Monthly Mean Land Surface Temperatures from MODIS Clear-Sky Retrievals. *J. Hydrometeorol.* 18, 2827–2847. https://doi.org/10.1175/JHM-D-17-0009.1

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell & S. Ermon (2016) Combining satellite imagery and machine learning to predict poverty. *Science,* 353**,** 790-4.

Project and related outputs (indicates published as a result of this award)**

Deng, G., Y. Liu, Y. Sun, and D.S. Matteson. " Designing Index-Based Insurance Using Novel Remote-Sensing Data." (2019). Working Paper.

**Fan, Y., Wen, J., Kira, O., McDonald, A., Liu, Y., Warner, J., Sun, Y., Regional scale crop yield estimation using satellite solar-induced chlorophyll fluorescence (SIF), *submission imminent.*

**Hu, L., Sun, Y., Collins, G., Fu, P. (2020), Improved Estimates of Monthly Land Surface Temperature from MODIS using a Diurnal Temperature Cycle (DTC) Model, *ISPRS Journal of Photogrammetry and Remote Sensing, 168, 131-140, (https://doi.org/10.1016/j.isprsjprs.2020.08.007)*

**Hu, L., Improvement of clear-sky LST monthly products by using Diurnal Temperature Cycle model (DTC), Jan. 12-15, 2020, Boston, MA, 2020 AMS conference.

**Hu, L. (2020), Better Measures of Land Surface Temperatures Improve Monitoring for Drought, Stressed Crops, and Crop Productivity. (https://www.agrilinks.org/post/lst).

**Kira, O., and Sun, Y. (2020), Extraction of sub-pixel C3/C4 emissions of solar-induced chlorophyll fluorescence (SIF) using artificial neural network, *ISPRS J PHOTOGRAMM.* doi.org/10.1016/j.isprsjprs.2020.01.017

**Sun, Y. (2020), Measuring Plant Photosynthesis from Space to Understand Crop Production on the Ground: The Promise of Solar-Induced Chlorophyll Fluorescence (https://www.agrilinks.org/post/measuring-plant-photosynthesis-space-understand-crop-production-ground-promise-solar-induced).

Tang, B., Y. Sun, Y. Liu, and D.S. Matteson. "Dynamic Poverty Prediction with Vegetation Index." (2018). Workshop on Modeling and Decision-Making in the Spatiotemporal

Domain, 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

**Wen, J., P. Köhler, G. Duveiller, N.C. Parazoo, L. Yu, C. Y. Chang, and Sun, Y. (2020), A Framework for Harmonizing Multiple Satellite Instruments to Generate a Long-Term Global High Spatial-Resolution Solar-Induced Chlorophyll Fluorescence (SIF), *Remote Sens. Environ.* doi.org/10.1016/j.rse.2020.111644. Datasets available at: https://www.yingsun.info/datasets-and-tools.

Yu, L., J. Wen, C.Y., Chang, C. Frankenberg, Y. Sun (2019), High Resolution Global Contiguous Solar-Induced Chlorophyll Fluorescence (SIF) of Orbiting Carbon Observatory-2 (OCO-2), *Geophysical Research Letters* 46(3): 1449-1458. *https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018GL081109.* Datasets archived at Oak Ridge National Lab (ORNL) DACC https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1696