

Learning to predict crop type from heterogeneous sparse labels using meta-learning

Gabriel Tseng
NASA Harvest

gabrieltseng95@gmail.com

Catherine Nakalembe
University of Maryland, College Park

cnakalem@umd.edu

Hannah Kerner

University of Maryland, College Park

hkerner@umd.edu

Inbal Becker-Reshef

University of Maryland, College Park

ireshef@umd.edu

Abstract

There are many labelled datasets relating to land cover and crop type mapping that cover diverse geographies, agroecologies and land uses. However, these labels are often extremely sparse, particularly in low- and middle-income regions, with as few as tens of examples for certain crop types. This makes it challenging to train supervised machine learning models to detect specific crops in satellite observations of these regions. We investigate the utility of model-agnostic meta-learning (MAML) to learn from diverse global datasets and improve performance in data-sparse regions. We find that in a variety of countries (Togo, Kenya and Brazil) and across a variety of tasks (crop type mapping, crop vs. non-crop mapping), MAML improves performance compared to pretrained and random initial weights. We also investigate the utility of MAML for different target data-size regimes. We find MAML outperforms other methods for a wide range of training set sizes and positive to negative label ratios, indicating its general suitability for land use and crop type mapping.

1. Introduction

Although remote sensing has plentiful unlabelled data, collecting high-quality labels is much more challenging. While certain land-cover (LC) classes can be labelled using satellite photo-interpretation, crop type labels typically require teams to be sent into the field, which is expensive and not always feasible. This limits the scale of in-situ data collection efforts that can be undertaken. In some areas, it may not be possible to collect additional labels (e.g., due to conflict), making it especially important to develop methods that can efficiently learn from small datasets.

While extensive data collection efforts are rare, there

are many small-scale efforts to collect LC and crop type datasets globally. These datasets are extremely heterogeneous, capturing different information depending on their geographic coverage or the project's scope and aims. Efforts to learn from such heterogeneous datasets typically require careful curation and pipeline design depending on the end goal [38, 40, 16], or reduce the granularity of the dataset so that the labels are homogeneous (e.g., reducing crop type labels to coarser crop/non-crop labels) [44], in which case some information in the labels is lost.

In this paper, we demonstrate the applicability of meta-learning (specifically model-agnostic meta-learning, or MAML [10]) to learning from many heterogeneous datasets without losing the unique information contained in each dataset, and apply a MAML formulation that can be used for a variety of LC and crop type classification scenarios. We investigate the effectiveness of this model in a variety of geographies and across different sample-size regimes, and apply it to create a common bean map for Busia province, Kenya and a coffee map in Luís Eduardo Magalhães municipality, Brazil given very few positive task labels (42 and 20, respectively). All source code¹, photo-interpreted labels² and maps³ from this study are made publicly available to promote operational uptake and future research.

2. Related Work

Machine learning has commonly been used to identify cropland from Earth observation data. Common methods include tree-based classifiers [37, 24, 27] or deep neural networks [22, 21]. However, these have often been applied to small, homogeneous regions [22, 21] or require

¹<https://github.com/nasaharvest/crop-maml>

²<https://doi.org/10.5281/zenodo.4680394>

³<https://code.earthengine.google.com/39a0fedfc7ac7f21c3dcb06eab29917d>

large, task-specific datasets (hundreds of thousands of labels) [30, 42, 29, 27].

A number of techniques have been developed to improve performance on a data-sparse task by leveraging a larger dataset of related samples. In remote sensing, the following approaches have been applied:

Transfer Learning Transfer learning consists of first training a model to perform well on an initial source dataset before finetuning it on the dataset of interest. Transfer learning works best with extremely large source datasets such as the ImageNet dataset [28]. Pretrained models are typically optimized for a single target task, such as aircraft detection [7] or land cover classification [38]. Alternative pretraining datasets can also be used—[43] used an unsupervised tree detection algorithm to create a large pretraining dataset for tree-crown detection. Finally, pretraining a model on one geography before applying it to another has been used for crop yield prediction [41], but this requires very similar tasks across the different regions.

Multi-task learning Multi-task learning consists of solving multiple tasks at the same time [5], for example, simultaneously predicting tree type and tree health from satellite imagery [6]. In addition, multiple classification tasks can be learned together, e.g., [20] used multi-task learning for binary cropland (crop vs. non-crop) mapping, simultaneously training a model to classify pixels from a global crop/non-crop dataset and a smaller, region-specific crop/non-crop dataset.

Meta-learning Meta-learning consists of using a variety of related tasks to learn how to efficiently learn when given new tasks. Model-agnostic meta-learning [10] consists of learning an initial set of model weights θ that are optimized to efficiently learn new tasks. This is in contrast to transfer learning in which the initial weights θ are optimized by pretraining the model on a larger dataset. Compared to pretraining, meta-learning enables efficient learning of information in classes with very few labels, as each of these classes can be a single task. In pretraining, the underrepresentation of these classes can make it much harder for the model to learn from them [15, 8].

Rußwurm et al. [31] applied model-agnostic meta-learning (MAML) [10] to land cover classification. Specifically, they defined each task to be land cover classification within a single biosphere and trained a convolutional network on the SEN12MS segmentation dataset [34] and the DeepGlobe dataset [9].

Our work builds on Rußwurm et al.’s initial exploration into meta-learning for cross geography generalization. We investigated the ability of MAML models to effectively learn from heterogeneous sparse datasets, reformulating the

task so that this method can be applied to classes, such as specific crops, that are not present in all tasks. In addition, we investigated the performance of the model in a variety of real-world settings, using this method to generate crop type maps in Brazil and Kenya given very few labels for the target task.

3. Data

3.1. Labelled Data

We leveraged multiple datasets to construct the tasks for meta-learning. In total, we assembled 50,169 samples, the majority (41,364 or 82%) of which were binary crop/non-crop labels. We additionally had access to 8,805 crop type labels spanning 18 crop type classes (including monocropped and intercropped, in which only one or more than one crop is planted in the same field respectively), 4 land use labels and 4 countries (Brazil, Kenya, Togo and Mali). The distribution of labels is shown in Figure 1. We describe these datasets below.

GeoWiki global crowd-sourced labels We leveraged a large (35,866 samples) publicly-available crowd-sourced dataset of binary cropland (crop/non-crop) labels from diverse, globally-distributed locations from GeoWiki (<https://www.geo-wiki.org/>). The sampling, labelling and quality assessment procedures are described in Bayas et al. [4]. As each point was labelled by multiple labellers, we took the mean of all labels for each point and used a threshold of 0.5 to convert the mean to a binary label.

Groundtruth crop type labels We obtained 3 datasets with crop type labels collected in the field in 3 countries: Brazil, Mali and Kenya. Specifically, we used the following datasets:

- **Kenya:** Local labels from Kenya were obtained from in-country partners who work with field agents. These labels include agricultural fields collected by Plant Village [17, 18]. The dataset contains 8,318 total labels and 13 classes. These included 12 single-crop classes: cassava (365 labels), common bean (42 labels), cowpea (41 labels), green grams (127 labels), groundnut (3 labels), maize (830 labels), millet (25 labels), sorghum (182 labels), soybean (6 labels), sugarcane (7 labels), sunflower (60 labels), and wheat (12 labels). We aggregated all fields that contained more than one crop type into one class of intercropped fields (6,227 labels).
- **Mali:** This dataset was collected under the **Relief to Resilience in the Sahel (R2R)** project to gather valuable on-the-ground information about crop conditions for relevant government agencies in Mali

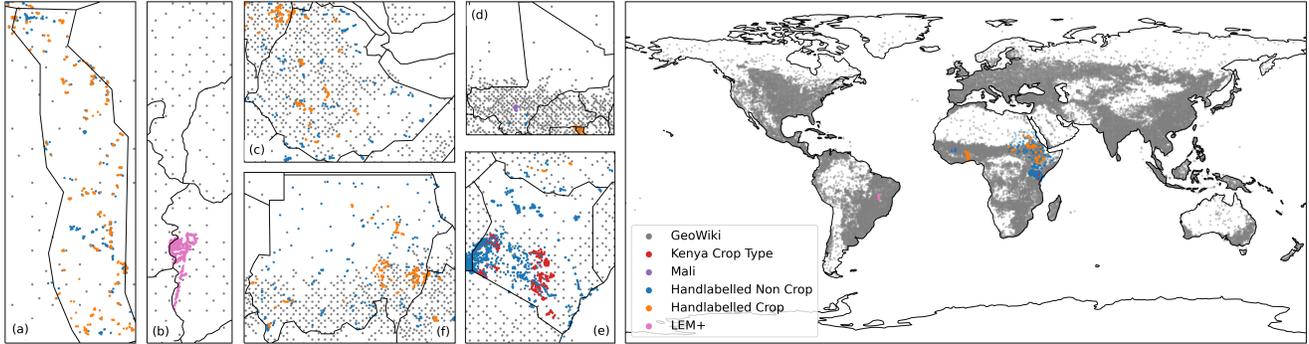


Figure 1: The spatial distribution of the labelled datasets. The insets show, clockwise from left, (a) Togo, (b) central Brazil, including state boundaries, (c) Ethiopia, (d) Mali, (e) Kenya and (f) Sudan, highlighting the combinations of labelled datasets in each country.

in partnership with NASA Harvest (<https://nasaharvest.org/>). This dataset consists of 4 crops in Segou, Mali: maize (35 labels), sorghum (44 labels), millet (55 labels) and rice (14 labels).

- **LEM+ (Brazil):** We used the open source LEM+[23] dataset which labeled monthly land use in 1,854 fields across 16 land use classes between October 2019 and September 2020. We took all fields that had the same land use class between October 2019 and April 2020, yielding 7 classes: cerrado (or natural vegetation) (149 labels), pasture (95 labels), coffee (20 labels), hay (21 labels), conversion area (recently deforested area that was not previously cerrado) (11 labels), eucalyptus (26 labels) and bracharia (6 labels).

Hand-labelled data: While crop type labels usually require groundtruth observation in the field, in most cases binary cropland labels can be determined based on visual interpretation of high-resolution satellite images. We used the following datasets collected in this way:

- **Togo:** We used a dataset of 1,319 crop/non-crop labels in Togo provided by [20]. These points were labeled based on expert interpretation of a high-resolution (< 1m/pixel) SkySat basemap of Togo from 2019 [36]. In addition, we used the provided test set of 350 randomly sampled points in Togo, labelled by consensus from 4 expert labellers.
- **Crop:** We used crop-labels collected in Sudan (289 labels) and Ethiopia (454 labels), labeled based on expert interpretation of high resolution (3 m/pixel) PlanetScope basemaps [36]
- **Non-Crop:** We supplemented the groundtruth and handlabelled crop datasets with handlabelled non-crop

points, labeled by experts based on visual interpretation of PlanetScope basemaps [36]. We generated 2,697 non-crop labels in Kenya, 142 non-crop labels in Mali, 202 non-crop labels in Ethiopia and 376 non-crop labels in Sudan.

3.2. Satellite data

We used Sentinel-2 top-of-atmosphere reflectance (Level 1C) observations from Google Earth Engine (GEE) as input to the model (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2). To construct a cloud-free representation of the pixel, we used the algorithm in [33] to find the least-cloudy pixel within a 30-day time period, resulting in a 12 month least-cloudy time series for each label. We used all Sentinel-2 bands except B1 (coastal aerosols) and B10 (cirrus SWIR). We additionally included the normalized difference vegetation index (NDVI) ($NDVI = \frac{B08 - B04}{B08 + B04}$). All bands (ranging from 10m to 60m resolution) were upsampled to 10m during the GEE export.

To construct the 12-month time series we used observations acquired between March of Year N and March of Year N+1 where N is the year the labels were created for. For example, the GeoWiki labels were based on satellite images from 2017 [35], thus we used observations acquired March 2017-March 2018 for those samples. Our handlabelled datasets were based on images from 2019, thus the time series spans March 2019-March 2020. The labels for all other datasets were acquired between 2018-2020; each groundtruth label has a date attribute that we used to determine the time series start and end dates. Obtaining the correct time window is important, as land use can change over time [23], leading to incorrect labels if the wrong time period is paired with a label.

For each label, we exported a $160m \times 160m$ (16×16 Sentinel-2 pixels) patch around the label location using

GEE. This was necessary to give sufficient spatial context to the cloud filtering algorithm. We then took the closest pixel to the label within the patch. We focused on modelling the temporal structure of the data instead of the spatial structure, as prior work has highlighted the importance of temporal structure in crop and crop type mapping [32, 19], and has shown that successful crop classification models use the majority of their parameters to model the temporal structure of the data [12].

4. Model Agnostic Meta-Learning

Algorithm 1 Model-Agnostic Meta-Learning

- 1: **Require:** $p(\mathcal{T})$: Distribution over tasks
 - 2: **Require:** α, β : step size hyperparameters
 - 3: randomly initialize θ
 - 4: **while** not done **do**
 - 5: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
 - 6: **for all** \mathcal{T}_i **do**
 - 7: Evaluate $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$ with respect to K examples
 - 8: Compute adapted parameters with gradient descent: $\theta'_i \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$
 - 9: Update $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$
-

Model-agnostic meta-learning (MAML) [10] is an algorithm for learning a set of initial model weights θ that are optimized to efficiently learn new tasks with few examples and/or gradient steps. The MAML model learns these initial weights θ by training on other tasks, such that these tasks behave as training examples (Algorithm 1).

Task construction We used two types of tasks to train our meta-model:

1. **Crop vs. non-crop tasks:** We combined all the datasets described in Section 3.1 and split the combined dataset along national boundaries (using boundaries from [26]). We omitted countries with fewer than 10 positive or negative examples. This resulted in 91 tasks for which the model learned to classify pixels as crop or non-crop.
2. **Crop type vs. rest tasks:** As described in Section 3.1, we had crop type labels for Brazil, Mali and Kenya. For each crop type C in each country, we constructed a binary “ C vs. rest” task. The “rest” class consisted of all other crop, crop type, and non-crop labels in the country. Differentiating a crop from other crops is substantially more challenging than differentiating a crop from non-crop pixels (Figure 2). To compensate for this, we upsampled the other crop samples by a factor of 10 so they would be over-represented in the negative samples for each task. To prevent overfitting, we

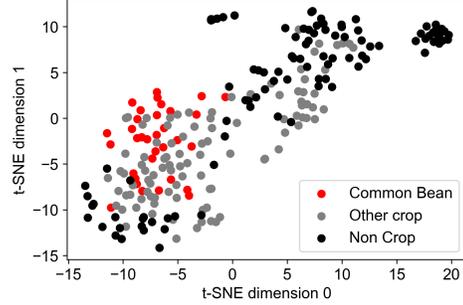


Figure 2: A t-SNE [39] plot of a random subset of the raw pixel-wise satellite time-series drawn from non-crop, common bean and other crop classes in Kenya. The crop and non-crop instances are clustered much more effectively than the common bean and other crop instances, highlighting the more challenging nature of the crop type vs. rest problem compared to crop vs. non-crop.

sampled each task-batch so that it would only contain unique instances.

Model architecture MAML can be applied to any neural network architecture. We used a one-layer LSTM (Figure 3) as in [20] to classify the time series for a single pixel as crop vs. non-crop or crop type vs. rest (depending on the task type). The LSTM had a hidden vector size of 128 and the final hidden output was passed to a 2-layer classifier and sigmoid activation. We applied variational dropout [11] between each LSTM timestep with 20% of weights randomly dropped (dropout value of 0.2).

Training procedure We trained the meta-model in a one-shot learning regime, using a binary cross entropy loss function. Each batch consisted of 10 positive and negative examples (20 total) from each task, sampled without replacement until all positive or negative examples from a task had been sampled, at which point all instances would be replaced. As in [1], we applied cosine annealing to the meta-learning rate (β) when training the model, with a maximum value of 10^{-4} and a minimum value of 10^{-6} . We used an update learning rate (α) of 10^{-4} . We sampled all tasks before updating the meta-weights, so that each batch included all tasks. We implemented the model using PyTorch [25] and learn2learn [2].

We randomly split the tasks into training and validation tasks using an 80/20 split. The test sets described in Section 5.1 were held out from both the training and validation sets. We selected the model that performed best on the validation tasks after 2000 passes over all tasks.

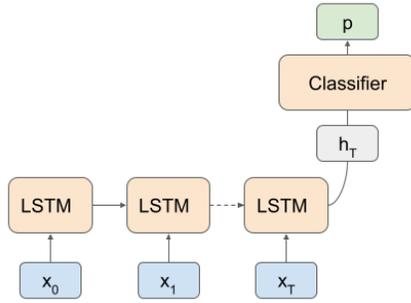


Figure 3: The LSTM model used to classify pixels as containing a crop (or the crop of interest) or not.

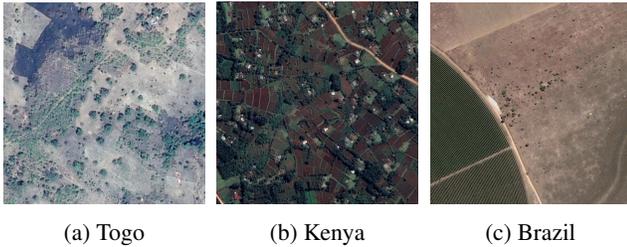


Figure 4: Example $500\text{m} \times 500\text{m}$ satellite images of the evaluation regions, demonstrating the variety in field sizes and agro-ecology being evaluated. (Images obtained from Google Earth Pro basemaps, comprised primarily of high resolution Maxar images.)

5. Experiments

5.1. Evaluation datasets

We evaluated the model on 3 datasets. Figure 4 illustrates the diversity of agro-ecology and cropping systems in each region.

1. **Togo:** We used the same dataset with crop vs. non-crop labels from [20]. This consisted of a training dataset with 1,319 hand-labelled examples (688 crop and 588 non-crop) as well as 43 examples in Togo from the GeoWiki dataset [35]. The test dataset contained 350 randomly sampled points within Togo with crop/non-crop examples labeled by consensus from four experts.
2. **Kenya:** We held out the 42 examples labeled as common bean in the Kenya crop type dataset to evaluate model performance for a minority class. We trained the model in a common bean vs. rest regime as described in Section 4.
3. **Brazil:** We held out the 20 coffee examples from the LEM+ dataset and trained the model in a coffee vs. rest regime as described in Section 4.

Each evaluation dataset consisted of a fine-tuning (training) and test dataset. For Togo, the test dataset consisted of the 350 randomly sampled test labels provided with the training dataset. For Kenya and Brazil, due to the small number of positive examples, we randomly sampled 10 positive and 10 negative instances from the evaluation set to create the test set. With very small sample sizes and the small evaluation sets used for Kenya and Brazil, the test performance can be sensitive to a particular split of the dataset and not representative of the performance on the dataset overall. To address this, we used bootstrapping to evaluate the model across different splits and report performance metrics as an average of 10 bootstraps.

Starting from initial weights learned using MAML, we finetuned the model on the evaluation dataset by training the model for 2000 iterations. Each iteration consists of a gradient update with batch size fixed to be the same size as during meta-training (10 positive and 10 negative examples). We sampled batches without replacement until all the instances in a class had been used, then replaced all instances. After finetuning, we evaluated the model on the held-out test set (or bootstrap sample), constructing the test set to only contain unique instances. We evaluated the model by measuring its Area Under the Receiver Operating Characteristic Curve (AUC ROC), which measures the probability the model will rank a randomly chosen positive instance above a randomly chosen negative instance (or how well the model separates positive and negative instances)[3]. We designed this evaluation procedure to emulate the common real-world scenario in which there is a need to create a crop type map for a region of interest but there are only a handful of groundtruth labels available for model training and evaluation.

5.2. Baselines

We compared our meta-learning method with two baseline methods for initializing the model weights prior to fine-tuning: 1) **Pretraining**, where the LSTM model is pre-trained to classify crop vs. non-crop using all available data, and 2) **Random** initialization, where the LSTM model weights are initialized using Xavier initialization [13].

In the case of the Togo dataset, we also compared our results to those from the multi-task learning (multi-headed LSTM) approach reported in [20].

6. Results

The overall results are shown in Table 1. Overall, across a variety of geographies and sample-size regimes, MAML outperformed both the random and pretrained baselines.

6.1. As a function of sample size

Using the Togo dataset, we compared the performance of the MAML model with the random and pretrained base-

Dataset	#Pos	#Neg	Model	AUC ROC
Kenya	32	32	MAML	0.849
			Pretrained	0.822
			Random	0.803
	4373	MAML	0.898	
		Pretrained	0.853	
		Random	0.870	
Brazil	20	20	MAML	0.979
			Pretrained	0.938
			Random	0.962
	420	MAML	0.998	
		Pretrained	0.974	
		Random	0.997	
Togo	10	10	MAML	0.844
			Pretrained	0.819
			Random	0.750
	688	588	MAML	0.910
			Pretrained	0.856
			Random	0.840
			Multi-headed[20]	0.894

Table 1: The AUC ROC results of crop type classification for the MAML, pretrained and randomly initialized models. The #Pos and #Neg values indicate the number of unique positive and negative instances used for training. All MAML, pretrained and random were obtained using 10 bootstraps. The multi-headed LSTM results were obtained using 3 bootstraps[20].

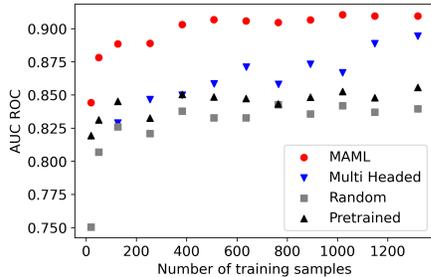


Figure 5: The effect of increasing the training sample size on model performance on the Togo test dataset. The results are taken at the 2000th training iteration. The multi-headed results are taken from [20]. The results from [20] are from 3 bootstraps, while the MAML results are from 10 bootstraps.

lines as the number of available training samples increases (Figure 5). We evaluated the models on the 350 randomly-sampled test instances. We additionally compared the model to the results reported for the multi-task model originally trained on the same training dataset[20]. Because the multi-headed LSTM was simultaneously trained on the GeoWiki [35] and Togo-specific dataset, its training procedure is different from the MAML model and the baselines. Most

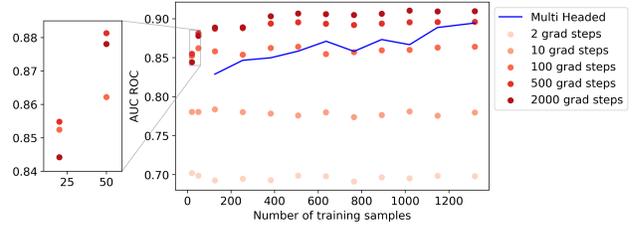


Figure 6: The performance of the MAML model on the Togo test dataset across different sample sizes as the number of finetuning gradient steps increases from 2 to 2000. The multi-headed LSTM results from [20] are included for reference. The results from [20] are from 3 bootstraps, while the MAML results are from 10 bootstraps.

notably, the training procedure in [20] used early stopping and a validation set instead of training for a fixed number of gradient steps. Overall, the MAML model outperformed all compared methods at all sample sizes.

In addition, the performance of the MAML model rapidly approaches the performance of the multi-headed LSTM [20] trained on all available data. When trained on only 254 samples, the MAML model achieved an average AUC ROC of 0.889, compared to an AUC ROC of 0.894 for the multi-headed LSTM trained on all the data (1,319 samples).

6.1.1 As a function of gradient steps

With small training set sizes, there may not be enough data samples to create a validation set that can be used to determine when to stop training (or other hyperparameter settings). To better understand the effect this may have on model over-fitting, we measured the performance of the MAML model on the Togo test set at a variety of training iterations (Figure 6).

With 20 data samples (10 positive and 10 negative), the model is sensitive to overfitting. Specifically, the model’s AUC ROC decreases from a peak of 0.855 at 500 gradient steps to 0.844 at 2000 gradient steps. However, for larger sample sizes the model is robust to many gradient steps, with model performance remaining stable from 500 to 2000 gradient steps.

6.2. As a function of class imbalance

In the cases of the Kenya and Brazil evaluation tasks (and other field data collection scenarios), we are limited by the number of positive groundtruth (crop type) examples that were collected in the field. However, we can more easily collect additional negative non-crop labels by visually interpreting and annotating satellite images. The downside is that adding more negative non-crop labels will increase the class imbalance. To evaluate the trade off between class

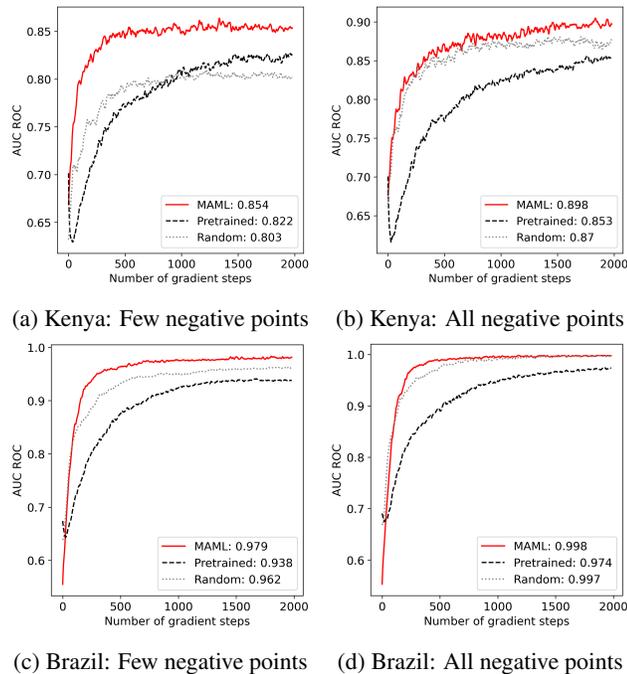


Figure 7: The averaged test AUC ROC of the MAML, random and pretrained models for the Kenya common beans and Brazil coffee test sets using (a, c) all positive samples and an equal number of positive samples, and (b, d) all available negative samples, across 10 bootstraps. This performance is plotted as a function of the number of training gradient steps. The values in the legend indicate the final AUC ROC of each model.

imbalance and plentiful, easy to collect negative labels, we tested the model (a) in a balanced regime, with all available positive samples and an equal number of negative samples, and (b) in an imbalanced regime, with all available positive samples and all available negative samples.

We found that even when using all available negative samples, the MAML model outperformed the random and pre-trained models (Figure 7). In addition, as in section 6.1.1 we found that the models were robust to overfitting in both in the balanced and imbalanced regimes, with model performance increasing or remaining stable as the number of gradient steps increased.

In addition, we found that in all cases except one (the balanced Kenya task), pretraining the model on crop vs. non-crop failed to improve the performance of the model compared to a random initialization. There is evidence that given the right training procedure, randomly initialized models can match the performance of those trained on ImageNet, even with tens of thousands of samples [14], but there are also many cases where pretraining does improve model performance [38, 7, 41]. We hypothesize that train-

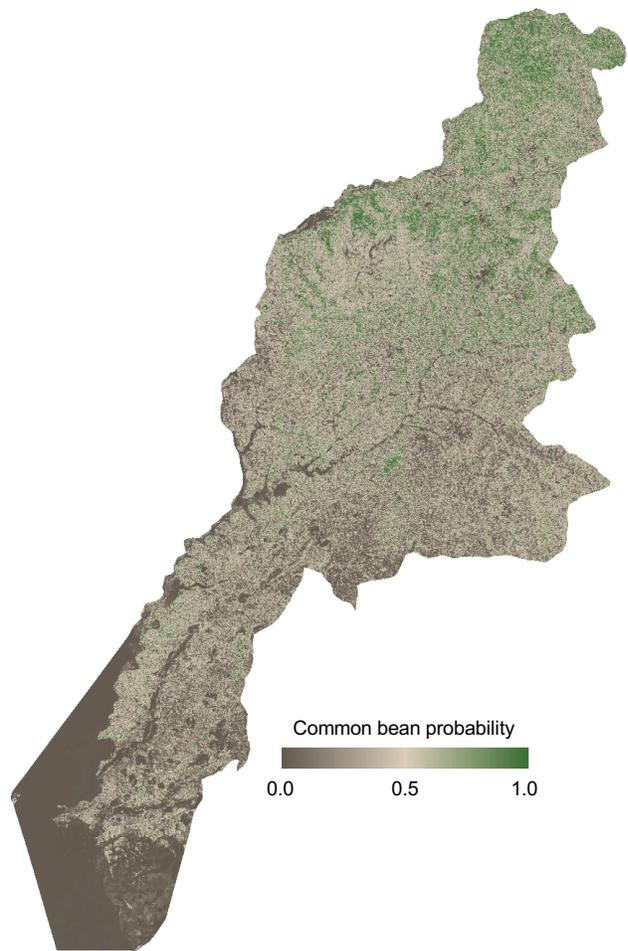


Figure 8: Common beans vs. rest map for the 2019-2020 season in Busia, Kenya, created from an ensemble of 10 bootstrapped MAML models.

ing the pretrained model on a crop vs. non-crop regime makes it more difficult for the model to learn to differentiate a single crop from other crops, leading to the lower performance. This highlights the suitability of pretraining when there is alignment between the source and target tasks, but that the problems may need to be carefully constructed to ensure pretraining benefits the test task [31]. An additional benefit of the MAML model is that it can be re-purposed for a variety of different tasks (e.g., crop mapping in Togo or crop type mapping in Brazil and Kenya) with no alterations.

6.3. Crop type maps from ensembles

We ensembled the the 10 bootstrapped crop vs. rest models to create a common bean map for Busia, Kenya (Figure 8) and a coffee map for Luís Eduardo Magalhães municipality in Brazil (Figure 9). We did this by using the trained

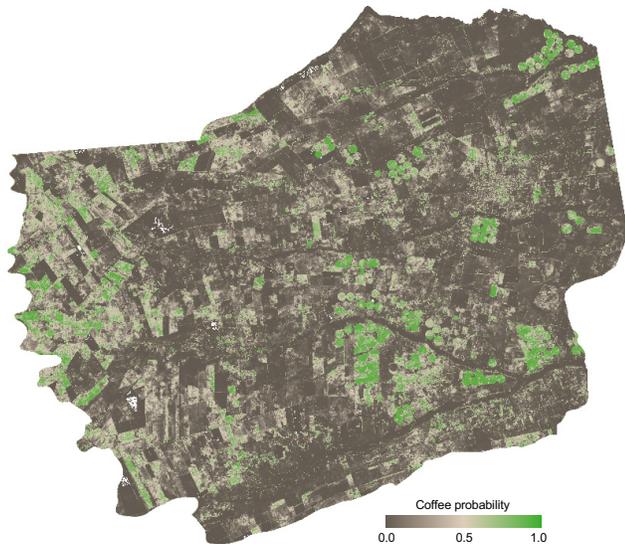


Figure 9: Coffee vs. rest map for the 2019-2020 season in Luís Eduardo Magalhães, Brazil, created from an ensemble of 10 bootstrapped MAML models.

models to create predictions for the entire area, and taking the mean of the predictions of the 10 models.

Given the sparsity of the labels, evaluating the maps is challenging. The LEM+ [23] dataset collected polygons delineating the boundaries of entire fields in Luís Eduardo Magalhães municipality, so we can measure the accuracy of the ensemble wall to wall crop type map by comparing it to these field polygons. We emphasize that these samples have some overlap with the training dataset, because we used the central pixels of each fields for training. In addition, these fields do not completely cover the municipality. Still, this metric is useful to understand how well the model can generate wall-to-wall (dense) maps given few sparse labels. Using a threshold of 0.5 to classify pixels as containing coffee or not, this map accurately classifies 87.5% of pixels in coffee fields, and misclassifies 2.8% of pixels in non-coffee fields as coffee. We computed intersection over union (IOU) by removing all pixels not covered by the LEM+ polygons, and calculating IOU using the remaining pixels. The IOU was measured to be 0.51.

7. Conclusions and Future Work

In conclusion, we presented a framework for combining multiple sparse, heterogenous datasets using model-agnostic meta-learning. We compared this approach against pretrained and randomly initialized weights in a variety of agro-ecologies, geographies and data regimes. We found that meta-learning is an effective way of combining many diverse datasets without losing granular information, and that this method outperformed the pretraining and random-

weight baselines for both crop vs. non-crop tasks and crop type vs. rest tasks.

In addition, we used this method to generate a common bean map for Busia province, Kenya and a coffee map for Luís Eduardo Magalhães municipality, Brazil, demonstrating the operational utility of this method for creating crop type maps given few labels.

In future work, we plan to investigate methods of communicating task-specific information (such as the agro-ecology of the region being learned) to the model and to continue extracting information from the labels to improve the model output (e.g., whether a field is monocropped or intercropped). In addition, we plan on expanding this method, both by adding additional training tasks (and extending the crop type tasks to include one v. one crop type tasks) and by using it to generate other crop type maps around the world.

References

- [1] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your MAML. In *International Conference on Learning Representations*, 2019. 4
- [2] Sébastien M R Arnold, Praateek Mahajan, Debajyoti Datta, Ian Bunner, and Konstantinos Saitas Zarkias. learn2learn: A library for Meta-Learning research. Aug. 2020. 4
- [3] Marina Barsky. Lecture notes in cscsi 485: Data mining 2012, 2012. 5
- [4] Juan Carlos Laso Bayas, Myroslava Lesiv, François Waldner, Anne Schucknecht, Martina Duerauer, Linda See, Stefan Fritz, Dilek Fraisl, Inian Moorthy, Ian McCallum, et al. A global reference database of crowdsourced cropland data collected using the geo-wiki platform. *Scientific data*, 4:170136, 2017. 2
- [5] Rich Caruana. *Multitask Learning*, page 95–133. Kluwer Academic Publishers, USA, 1998. 2
- [6] Tony Chang, Brandon P. Rasmussen, Brett G. Dickson, and Luke J. Zachmann. Chimera: A multi-task recurrent convolutional neural network for forest classification and structural estimation. *Remote Sensing*, 11(7), 2019. 2
- [7] Zhong Chen, Ting Zhang, and Chao Ouyang. End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sensing*, 10(1), 2018. 2, 7
- [8] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9268–9277, 2019. 2
- [9] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018. 2
- [10] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks.

- In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. 1, 2, 4
- [11] Yarın Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. 4
- [12] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata. Time-space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 6247–6250, 2019. 4
- [13] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterton, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. JMLR Workshop and Conference Proceedings. 5
- [14] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927, 2019. 7
- [15] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6:429–449, 2002. 5. 2
- [16] Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016. 1
- [17] Annalyse Kehs, Peter McCloskey, John Chelal, Derek Morr, Stellah Amakove, Bismark Plimo, John Mayieka, Gladys Ntango, Kelvin Nyongesa, Lawrence Pamba, Melodine Jep-too, James Mugo, Mercyline Tsuma, Winnie Onyango, and David Hughes. From village to globe: A dynamic real-time map of african fields through plantvillage. *bioRxiv*, 2019. 2
- [18] Annalyse Kehs, Peter McCloskey, and David Hughes. Busia county, kenya agricultural fields may 2019. <https://scholarsphere.psu.edu/resources/47e44778-5238-400a-af3e-a44abcc3c0d2>. 2
- [19] Hannah Kerner, Ritvik Sahajpal, Sergii Skakun, Inbal Becker-Reshef, Brian Barker, Mehdi Hosseini, Estefania Puricelli, and Patrick Gray. Resilient in-season crop type classification in multispectral satellite observations using growth stage normalization. In *ACM SIGKDD Conference on Data Mining and Knowledge Discovery Workshops*, 2020. 4
- [20] Hannah Kerner, Gabriel Tseng, Inbal Becker-Reshef, Catherine Nakalembe, Brian Barker, Blake Munshell, Madhava Paliyam, and Mehdi Hosseini. Rapid response crop maps in data sparse regions. In *ACM SIGKDD Conference on Data Mining and Knowledge Discovery Workshops*, 2020. 2, 3, 4, 5, 6
- [21] Dinh Ho Tong Minh, Dino Ienco, Raffaele Gaetano, Nathalie Lalonde, Emile Ndikumana, Faycal Osman, and Pierre Maurel. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal sar sentinel-1. *IEEE Geoscience and Remote Sensing Letters*, 15(3):464–468, 2018. 1
- [22] Emile Ndikumana, Nicolas Baghdadi, Dominique Courault, Laure Hossard, and Dinh Ho Tong Minh. Deep recurrent neural network for agricultural classification using multitemporal sar sentinel-1 for camargue, france. *Remote Sensing*, 10:1217, 08 2018. 1
- [23] Lucas Volochen Oldoni, Ieda Del’Arco Sanches, Michelle Cristina A. Picoli, Renan Moreira Covre, and José Guilherme Fronza. Lem+ dataset: For agricultural remote sensing applications. *Data in Brief*, 33:106553, 2020. 3, 8
- [24] Adam Oliphant, Prasad Thenkabail, Pardhasaradhi Teluguntla, Jun Xiong, Murali Krishna Gumma, Russell Congalton, and Kamini Yadav. Mapping cropland extent of southeast and northeast asia using multi-year time-series landsat 30-m data using a random forest classifier on the google earth engine cloud. *International Journal of Applied Earth Observation and Geoinformation*, 81:110–124, 2019. 1
- [25] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 4
- [26] Tom Patterson and Nathaniel Vaughn Kelso. Natural earth. 4
- [27] Charlotte Pelletier, Silvia Valero, Jordi Inglada, Nicolas Champion, and Gérard Dedieu. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187:156–168, 2016. 1, 2
- [28] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 2
- [29] Marc Rußwurm and Marco Korner. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 2
- [30] Marc Rußwurm, Charlotte Pelletier, Maximilian Zollner, Sébastien Lefèvre, and Marco Körner. Breizhcrops: A time series dataset for crop type mapping. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS (2020)*, 2020. 2
- [31] Marc Rußwurm, Sherrie Wang, Marco Korner, and David Lobell. Meta-learning for few-shot land cover classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 200–201, 2020. 2, 7

- [32] M. Rußwurm and M. Körner. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1496–1504, 2017. [4](#)
- [33] Michael Schmitt, Lloyd Hughes, Chunping Qiu, and Xiao Zhu. Aggregating cloud-free sentinel-2 images with google earth engine. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7:145–152, 09 2019. [3](#)
- [34] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu. Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7:153–160, 2019. [2](#)
- [35] Linda See. A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform, 2017. [3](#), [5](#), [6](#)
- [36] Planet Team. Planet application program interface: In space for life on earth, 2018–. [3](#)
- [37] Pardhasaradhi Teluguntla, Prasad Thenkabail, Adam Oliphant, Jun Xiong, Murali Krishna Gumma, Russell Congalton, Kamini Yadav, and Alfredo Huete. A 30-m landsat-derived cropland extent product of australia and china using random forest machine learning algorithm on google earth engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144:325–340, 2018. [1](#)
- [38] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment*, 237:111322, 2020. [1](#), [2](#), [7](#)
- [39] L.J.P. van der Maaten and G.E. Hinton. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9(nov):2579–2605, 2008. Pagination: 27. [4](#)
- [40] Anna X. Wang, Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon. Deep transfer learning for crop yield prediction with remote sensing data. COMPASS ’18, New York, NY, USA, 2018. Association for Computing Machinery. [1](#)
- [41] Anna X. Wang, Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon. Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS ’18, New York, NY, USA, 2018. Association for Computing Machinery. [2](#), [7](#)
- [42] Sherrie Wang, William Chen, Sang Michael Xie, George Azzari, and David Lobell. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2):207, 2020. [2](#)
- [43] Ben G. Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11), 2019. [2](#)
- [44] Jun Xiong, Prasad Thenkabail, J Tilton, Murali Krishna Gumma, Pardhasaradhi Teluguntla, Russell Congalton, Kamini Yadav, J Dungan, Adam Oliphant, J Poehnelt, C Smith, and R Massey. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad) cropland extent 2015 africa 30 m v001, 2017. [1](#)