# Mitigating Low Agricultural Productivity of Smallholder Farms in Africa: Time-Series Forecasting for Environmental Stressors

**Maryam Tabar Dongwon Lee David P. Hughes Amulya Yadav**

The Pennsylvania State University, University Park, PA, USA

{mfg5544, dongwon, dhughes, amulya}@psu.edu

## Abstract

African smallholder farmers have struggled with low agricultural productivity for decades, partly due to their inability to proactively assess irrigation needs in their farms in the face of long-term climate change. In this paper, we tackle this challenge by employing data-driven techniques to develop forecasting tools for three widely used crop-productivity related variables (i.e., actual evapotranspiration, reference evapotranspiration, and net primary production), which can then be used by farmers to take corrective actions on their farms. Prior work in this domain, despite using data-driven methods, suffers from two major limitations: (i) they mainly focus on estimating variable values (as opposed to forecasting the future); and (ii) they mostly use classical Machine Learning (ML) prediction models, despite the abundance of data sufficient to train sophisticated deep learning models. To fill this research gap, we collaborate with PlantVillage, the world's leading non-profit agricultural knowledge delivery platform for African farmers, to identify ∼2,200 smallholder farm locations, and gather remote-sensed data of these farms over a period of five years. Next, we propose CLIMATES, a meta-algorithm leveraging structural insights about temporal patterns of this time-series data to accurately forecast their future values. We conduct extensive experiments to evaluate its performance in this domain. Our experimental results show that CLIMATES outperforms several state-of-the-art time-series forecasting models. We also provide insights about the poor performance of some competing models. Our work is being evaluated by officials at PlantVillage for potential future deployment as an early warning system in East Africa. We release the code at https://github.com/maryam-tabar/CLIMATES.

## Introduction

Smallholder farms (less than two hectares in size) and their farmers form the backbone of African agriculture and food security, and constitute a significant proportion of the Gross Domestic Product (GDP) of several African countries. For example, agriculture on smallholder farms is the primary means of livelihood for more than 60% people in Sub-Saharan Africa, and is responsible for ∼75% of the region's total agricultural production (Gollin 2014; Salami, Kamara, and Brixiova 2010). In addition, smallholder agriculture also plays a critical role towards meeting several Sustainable Development Goals (SDGs)[1] laid out by the United Nations, such as "no poverty and zero hunger". Thus, developing techniques to improve the productivity and profitability of smallholder farms in Africa is of critical importance, as it could lead to significant improvements in the well-being of many disadvantaged communities in Africa.

Unfortunately, increasing the productivity/profitability of smallholder agriculture is a challenging problem because of several reasons: (i) smallholder farmers find it difficult to protect their farms against biotic stressors (e.g., pest and disease outbreaks); (ii) they lack awareness about modern agricultural practices; and, most importantly, (iii) over the last few decades, climate change on the African continent has significantly hampered the ability of smallholder farmers to achieve high agricultural productivity (Harvey et al. 2014). In fact, the high reliance of smallholder farmers on rain-fed agriculture, coupled with a lack of knowledge about future climatic conditions result in highly uncertain situations for farmers. For example, farmers do not know the irrigation needs of their crops at any given point in time (Shimeles, Verdier-Chouchane, and Boly 2018). This is one of the primary factors behind consistently low agricultural productivity among African smallholder farmers. As such, it is of great importance to help them get a better understanding of future climatic conditions on their farms, so that they can proactively assess and address their irrigation needs.

In this paper, we tackle this important problem by developing CLIMATES (**Cl**ustering **I**nitialized **M**eta **A**lgorithm for **T**ackling **E**nvironmental **S**tressors), a Machine Learning (ML) based predictive tool for farmers to forecast three important crop-productivity related variables: (i) actual evapotranspiration (AET); (ii) reference evapotranspiration (RET); and (iii) net primary production (NPP). Intuitively, both AET and RET measure the amount of water present in soil to support crop growth, whereas NPP measures the amount of crop growth that occurs inside a farm. Generating accurate predictions for these three variables can help smallholder farmers understand their irrigation needs better, e.g., if the AET forecast for a smallholder farm shows stress (i.e., the forecasted AET value is less than what is required for healthy crop growth), then a farmer can proactively start

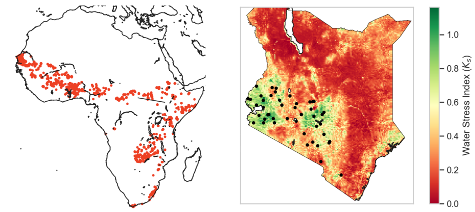


[1]https://www.un.org/sustainabledevelopment/

irrigating his/her farm to mitigate that stress.

To that end, this paper makes four novel contributions: (i) In collaboration with PlantVillage[2], we identify ∼2,200 smallholder farm locations across Africa, and gathered remote sensed data for AET, RET, and NPP for all these farm locations; (ii) We develop CLIMATES, an ML-based tool which leverages cluster-based structural insights of environmental time-series data in this domain, and then uses a distinct ML model to make (AET, RET, and NPP) forecasts for each cluster; (iii) We conduct a comprehensive analysis of the effectiveness of various popular classical ML and deep learning methods for time-series forecasting, and show that CLIMATES outperforms these state-of-the-art baseline models; and (iv) Finally, we provide insights about why generative models such as Variational Recurrent Neural Networks (VRNNs) (Chung et al. 2015), which explicitly model variability in sequential data, do not perform comparably.

Our work is done in collaboration with PlantVillage, and our results are being currently evaluated by them for potential real-world deployment as an early warning system for smallholder farmers in East Africa (who can use these warnings to proactively address their irrigation needs).




(a) Identified farm locations across Africa.

(b) A heatmap of water stress index ($K_s$) in Kenya.

Figure 1: Sample data of smallholder farms across Africa.

## Related Work

In this section, we discuss related studies in the agriculture and AI disciplines.

**Agriculture Research.** Numerous studies in the agriculture domain (Del Grosso et al. 2008; Sun and Du 2017; Zhang et al. 2017) have focused on estimating crop-productivity variables (i.e., AET, RET, and NPP) from meteorological factors, and finding associations between them. However, there have been a few attempts at using traditional models (such as KNN (Feng and Tian 2021) and ARIMA (Landeras, Ortiz-Barredo, and López 2009)) and neural models (Alves, Rolim, and Aparecido 2017) to predict ET. These studies found that statistical/ML models are more accurate than Historical Average methods (which do not involve learning). However, they reported mixed results when assessing the superiority of neural network models to other algorithms, e.g., Izadifar (2010) found that Multiple Linear Regression outperforms a neural network model in the task of predicting AET. However, this work has only considered Multi-Layer Perceptron as their neural network model, instead of using network architectures that were designed to explicitly model the sequential structure of time-series data, such as RNNs. In our work, we compare CLIMATES against much more stronger baselines such as VRNNs, Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997), etc.

**Artificial Intelligence Research.** To the best of our knowledge, there have been no prior studies in the AI community on forecasting these three crop-productivity variables across a large geographic region. However, there has been a large body of research on modeling sequential data for time-series forecasting. Some models, such as SARIMA (Hyndman and Khandakar 2008) and TBATS (Livera, Hyndman, and Snyder 2011), focused on explicitly modeling certain statistical properties of time-series data. Some other work in the neural network domain focused on tackling various challenges in different time-series data; e.g., State Frequency Model (SFM) combines the ideas behind LSTM and Discrete Fourier Transform to learn multiple frequency patterns from time-series data (Zhang, Aggarwal, and Qi 2017). In particular, one line of prior research focused on building deep latent variable models to capture variability in sequential datasets; for example, Jia et al. (2019) and Chien and Kuo (2017) proposed VRNN-based deep generative models for cropland detection and speech separation, respectively. In our work, we show that CLIMATES achieves higher predictive accuracy than many of these baselines.

## Dataset

Through our collaboration with PlantVillage, we identified 2,264 smallholder farm locations across Africa (as shown with red dots in Fig. 1a). For each farm location, we collected remote sensed time-series data for three variables (AET, RET, and NPP) over five years (from the beginning of 2015 to the end of 2019) from the WaPOR website[3], which is administered by the UN-FAO. For completeness, we provide a formal definition of these variables.

- **Actual Evapotranspiration (AET)**: AET refers to the summation of evaporation from soil, canopy transpiration, and interception. It can be used to derive the water demand of each crop; i.e., the difference between AET and RET (defined next) can be used to measure drought stress. Its unit is mm/day and its value ranges between 0.0 to 8.3 in our dataset (FAO 2018).
- **Reference Evapotranspiration (RET)**: RET refers to the evapotranspiration of a well-watered plant under well-defined standard conditions. Its unit is mm/day and its value ranges between 1.1 to 12.7 in our dataset (Allen et al. 1998; FAO 2018).
- **Net Primary Production (NPP)**: NPP refers to the amount of carbon dioxide absorbed by plants, and is an indicator of plant growth. The unit of NPP is gC/m$^2$/day (grams of carbon / square meter / day) and its value ranges between 0.0 to 9.265 in our dataset (FAO 2018).

**Data Characteristics.** The WaPOR website provides data for AET, RET, and NPP, with a spatial resolution of

[2]https://plantvillage.psu.edu

[3]https://wapor.apps.fao.org/home/WAPOR_2/1

0.00223° (∼250 m) and a temporal resolution of one dekad (∼10 days) (FAO 2018). Using this data, we generate three separate time-series datasets (one for each AET, RET, and NPP). Each dataset consists of 2,264 time-series data points (each data point is the time-series for a specific farm location), and the length of each time-series is 180 (since we collect dekadal data over five years, i.e., 36×5 = 180). For each dataset, we consider the first three years of data (i.e., from beginning of 2015 to end of 2017) as the training set. The data in 2018 (and 2019) is kept as the validation (and test) set, respectively. As a pre-processing step, we apply Min-Max normalization on the data of each farm, however, predictive performance metrics are computed after converting the data back to its original scale.

## The Meta-Algorithm: CLIMATES

In this section, first, we discuss key structural insights about our dataset which motivate the design of CLIMATES. Then, we describe our algorithm.

**Exploratory Data Analysis.** As shown in Fig. 1a, our 2,264 farm locations span widely across the African landmass. In total, these farm locations span across 20 different countries, each with its distinct climatic conditions. For example, while our farm locations in north-western Africa belong to the semi-arid Sahel region, farms in central Africa had tropical rain-forest climate, and farms in eastern and south-eastern Africa had savannah grassland climate, etc.

Due to this geographic and climatic diversity across our farm locations, we expect significant *variability* in all three of our datasets. To investigate this further, we cluster each dataset (separately) using an off-the-shelf feature-based clustering approach (Roelofsen 2018). At a high level, this clustering approach extracts the features of each time-series data point by applying Discrete Fourier Transform on its training portion. Once the feature vector for each time-series data point is extracted, bottom-up agglomerative clustering is used (with the Euclidean distance metric, and complete-linkage strategy for merging intermediate clusters).

As a result of data clustering, we obtain six clusters that have distinctly different shapes and patterns. Fig. 2 illustrates three of these clusters obtained on the NPP dataset (we see similar results on the AET and RET datasets). Due to this significant variability, therefore, *we hypothesize that forecasting methods that may work well for data points in one cluster may not necessarily work well on other clusters*. This crucial insight motivates our design of CLIMATES.

**The Proposed Meta-Algorithm.** Given this strong variability inside our datasets, we conducted a cluster-by-cluster comparison of the predictive performance of several popular classical and deep-learning based forecasting methods. This analysis would help us understand whether a single forecasting method works best across all clusters, or whether different methods work better in different clusters.

For this comparison, we consider the time-series data points belonging to each of our six clusters separately. Then, on the data of each cluster, we train and test a heterogeneous mix of statistical, classical ML, and deep learning methods, namely TBATS, SARIMA, Linear Regression (LR), Random Forest (RF) (Breiman 2001), XGBoost (Chen and Guestrin 2016), Support-Vector Machine (SVM) (Cortes and Vapnik 1995), LSTM, SFM, and Temporal Convolutional Network (TCN) (Bai, Kolter, and Koltun 2018).

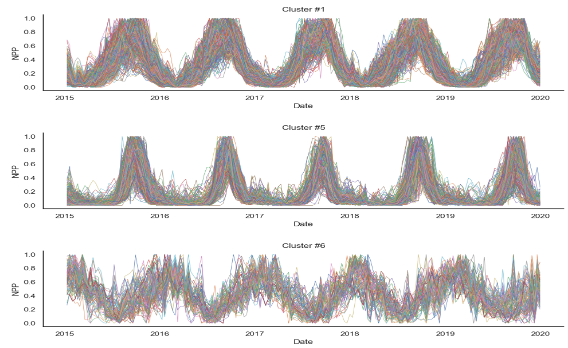



Figure 2: Three (out of six) NPP clusters generated through feature-based clustering

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| TBATS | 0.2110 | 0.2860 | 0.2491 | 0.5235 | 0.4417 | 0.2279 |
| SARIMA | 0.1896 | **0.2409** | **0.2112** | 0.3840 | 0.3518 | 0.2071 |
| LR | 0.1731 | 0.2424 | 0.2234 | 0.3891 | 0.3603 | 0.2127 |
| RF | **0.1726** | 0.2481 | 0.2159 | 0.3822 | 0.3510 | 0.2123 |
| XGBoost | 0.1807 | 0.2555 | 0.2313 | 0.3992 | 0.3788 | 0.2249 |
| SVM | 0.1889 | 0.2582 | 0.2486 | 0.3916 | 0.4165 | 0.2301 |
| LSTM | 0.1728 | 0.2505 | 0.2349 | **0.3774** | 0.3446 | **0.2035** |
| SFM | 0.1740 | 0.2446 | 0.2199 | 0.3800 | **0.3412** | 0.2186 |
| TCN | 0.1890 | 0.2618 | 0.2410 | 0.3817 | 0.3774 | 0.2099 |

Table 1: CV of different models on the NPP clusters.

Table 1 shows the coefficient of variation (CV)[4] achieved by the aforementioned methods on all six clusters found on the NPP dataset (analogous results on the AET and RET datasets are represented in Tables 2 and 3, respectively). Note that these results are for single-step forecasting, i.e., we try to predict the next dekadal NPP, AET, and RET values. These tables confirm that no single forecasting method works best across all clusters, e.g., on the NPP dataset, statistical methods like SARIMA work best on the second and third clusters, deep learning methods like LSTM and SFM work best on the fourth, fifth, and sixth clusters, whereas a Random Forest model works best on the first cluster. Thus, to get accurate forecasts consistently across the wide expanse of the African landmass, it is critically important to rely on an ensemble of well-trained models, each of which works well on a specific region of Africa.

Based on this finding, we now describe our meta-algorithm. CLIMATES works as follows: (i) It clusters the original time-series data using a feature-based clustering approach into different clusters. (ii) For each of these clustered datasets, it finds the best performing forecasting model (i.e., the model with lowest CV on the validation set of that cluster). We select the best performing model on each cluster out of the nine models shown in Table 1. Note that we use this selection of models inside CLIMATES to ensure a good het-

[4]Coefficient of variation (CV) refers to the root mean squared error divided by the average of the target variable. Therefore, the lower CV is, the better performance a method has.

|  | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| TBATS | 0.1947 | 0.2285 | 0.2380 | 0.3060 | 0.3564 | 0.2217 |
| SARIMA | **0.1713** | **0.2044** | 0.2197 | 0.2799 | 0.2990 | 0.2049 |
| LR | 0.1763 | 0.2051 | 0.2179 | 0.2806 | 0.2981 | 0.1976 |
| RF | 0.1725 | 0.2058 | 0.2180 | 0.2772 | 0.2834 | 0.1976 |
| XGBoost | 0.1742 | 0.2104 | 0.2187 | 0.2839 | 0.2943 | 0.2012 |
| SVM | 0.1769 | 0.2162 | 0.2262 | 0.3044 | 0.3145 | 0.2002 |
| LSTM | 0.1723 | 0.2097 | **0.2114** | **0.2669** | 0.2728 | 0.1984 |
| SFM | 0.1767 | 0.2115 | 0.2115 | 0.2709 | **0.2722** | **0.1967** |
| TCN | 0.1764 | 0.2058 | 0.2263 | 0.2693 | 0.2883 | 0.2016 |

Table 2: CV of different models on the AET clusters.

|  | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| TBATS | 0.1259 | 0.1435 | 0.1502 | 0.1029 | 0.0982 | 0.1663 |
| SARIMA | 0.1084 | 0.1261 | **0.1297** | 0.0910 | 0.0876 | 0.1552 |
| LR | 0.1014 | 0.1244 | 0.1509 | 0.0898 | 0.0885 | 0.1517 |
| RF | 0.0998 | 0.1233 | 0.1474 | 0.1050 | 0.0890 | 0.1496 |
| XGBoost | 0.1018 | 0.1229 | 0.1410 | 0.1038 | 0.0896 | 0.1515 |
| SVM | 0.1066 | 0.1307 | 0.1463 | 0.1148 | 0.0921 | 0.1508 |
| LSTM | **0.0988** | **0.1180** | 0.1372 | **0.0839** | 0.0820 | **0.1430** |
| SFM | 0.0991 | 0.1203 | 0.1346 | 0.0875 | **0.0792** | 0.1443 |
| TCN | 0.1022 | 0.1306 | 0.1365 | 0.0979 | 0.0850 | 0.1492 |

Table 3: CV of different models on the RET clusters.

erogeneous mix of statistical methods, classical ML methods, and deep learning methods. We further note that as more sophisticated time-series forecasting methods are developed, they can also be used as part of the CLIMATES ensemble. (iii) At test time, each time-series data point is assigned to a subset of clusters. We considered two general strategies for assigning data points to the clusters: (a) we assign each time-series data point to the nearest cluster (CLIMATES-I), (b) we assign each time-series data point to a subset of clusters that falls within $d$ distance from that data point. The threshold $d$ is set to two heuristically computed values: (1) the average distance between the data points and their closest cluster (CLIMATES-II), (2) the median distance between the data points and their closest cluster (CLIMATES-III). (iv) Finally, the best performing model on each selected cluster (in our chosen subset) is used to get a prediction on that test data point, and the average of the predicted values is returned as the final forecast of CLIMATES. We now conduct a rigorous evaluation of the predictive performance of our meta-algorithm against a comprehensive set of baselines.

## Experimental Evaluation

We provide two sets of results. First, we provide a brief background on the VRNN architecture and show results comparing the predictive performance of CLIMATES against VRNNs, which at least in theory, should serve as a strong baseline. Second, we show results comparing the predictive performance of CLIMATES against a wide variety of statistical/classical ML and deep learning models.

**VRNN Architecture.** VRNN is a deep generative model that extends the idea behind Variational Autoencoders (VAE) (Kingma and Welling 2014; Rezende, Mohamed, and Wierstra 2014) to sequential data. VRNNs can be viewed as a sequence of VAEs conditioned on the hidden state of an RNN. Thus, similar to VAEs, they consist of generative and inference networks; the latter encodes the input into a latent space, and the former generates the output by reconstructing the input from the latent space. In fact, the generative process at time $t$ begins with generating the latent variable $z_t$ from a Gaussian distribution. However, unlike VAE, $z_t$ is conditioned on $h_{t-1}$ (the hidden state of RNN at time $t-1$) to be able to model the consistency within a single time-series data point (Chung et al. 2015). During training, VRNN aims to maximize the log-likelihood of observations $\ell(p(x_{\leq T}))$, where $x_{\leq T} = \{x_1, ..., x_T\}$ represents the input time-series of length $T$. However, as inferring the log-likelihood is computationally intractable, VRNN maximizes the variational lower-bound of the log-likelihood given in Equation 1. This lower-bound consists of two terms: (i) reconstruction likelihood, and (ii) the KL distance between the approximate posterior and the prior distributions. In our paper, we compare CLIMATES against VRNNs because their ability to learn explicit representations of variability across time-series data points (through the sequence of latent variables $z_{\leq T}$) makes them ideal models for our domain.

$$\ell(p(x_{\leq T})) \geq \mathbb{E}_{q(z_{\leq T}|x_{\leq T})}[\sum_{t=1}^{T}(log(p(x_t|z_{\leq t}, x_{<t})) - KL(q(z_t|x_{\leq t}, z_{<t})||p(z_t|x_{<t}, z_{<t})))] \quad (1)$$

|  | AET | RET | NPP |
|---|---|---|---|
| CLIMATES-I | **0.2075** | **0.0989** | **0.2409** |
| $\text{VRNN}^{KL}_{Deterministic}$ | 0.2161 | 0.1052 | 0.2594 |
| $\text{VRNN}_{Deterministic}$ | 0.2166 | 0.1053 | 0.2639 |
| $\text{VRNN}_{Gaussian}$ | 0.2836 | 0.1504 | 0.4496 |
| $\text{LSTM}_{Deterministic}$ | 0.2113 | 0.1039 | 0.2617 |
| $\text{LSTM}_{Gaussian}$ | 0.2754 | 0.1507 | 0.3863 |

Table 4: A CV comparison between CLIMATES and VRNNs/LSTMs

**Comparison with VRNNs.** We now provide results comparing the performance of CLIMATES against VRNN and LSTM. In this set of experiments, a separate LSTM and VRNN is trained (and tested) on each of our three datasets. For both VRNN and LSTM, we experiment with two different output functions (Deterministic and Gaussian). Finally, the negative of the variational lower-bound given in Equation 1 is used as VRNN's loss function.

Table 4 compares the CV achieved by CLIMATES-I against VRNN and LSTM variants (for single-step forecasting on our three datasets). This table shows that regardless of the choice of output function, CLIMATES-I outperforms both VRNN and LSTM models. Surprisingly, CLIMATES-I, on average, achieves 6.3% lower CV than $\text{VRNN}_{Deterministic}$, even though VRNNs have latent variables to model variability inside our datasets. *Additionally, applying t-test, we find that the difference between the CV of CLIMATES-I and these models is statistically significant (p-value is consistently less than 0.01).* In fact, these results show that VRNN is unable to outperform LSTM on any dataset; in particular, $\text{VRNN}_{Deterministic}$ (which has stochastic latent states) cannot outperform $\text{LSTM}_{Deterministic}$ (which does not have any stochastic components). Counterintuitively, this shows that explicitly learning representations of variability inside our datasets does not seem to help.

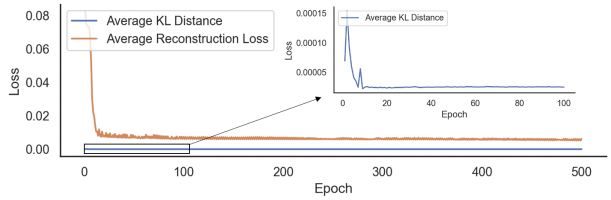



Figure 3: The learning curves of the two components of VRNN's loss function during training on the NPP dataset.

**Why Do VRNNs Not Work?** To understand VRNN's poor performance, we take a closer look at VRNN$_{Deterministic}$'s learning curves during training (we see similar results with other output functions). In particular, we separately analyze the learning curves of two components in the VRNN's loss function, i.e., (i) the reconstruction loss, and (ii) the KL term. Fig. 3 illustrates the changes in the values of these two components with increasing number of epochs on the NPP dataset. According to this figure, the KL distance vanishes into zero after a few epochs; i.e., the approximate posterior becomes equal to the prior in the early epochs, and hence, the model starts ignoring latent variables in the early stages of training (we see similar results on the AET and RET datasets). Thus, we observe that, in practice, training VRNN leads to a local optimum which hinders capturing variability across data points in a dataset, even though, in theory, it has the capability of capturing variations. Similar findings have been reported with VAEs, e.g., prior research found that the same issue (called *"posterior collapse"*) occurs in VAE (He et al. 2019). However, to the best of our knowledge, our work is the first to report this posterior collapse issue with VRNNs. Further, prior work proposed KL-annealing to tackle posterior collapse in VAEs (Bowman et al. 2016); however, as the second row of Table 4 (i.e., VRNN$^{KL}_{Deterministic}$) shows, even with KL-annealing, VRNNs are unable to beat CLIMATES-I.

**Comparison with Other Baselines.** Having established the superior predictive performance of CLIMATES over VRNNs and LSTMs in Table 4, we now evaluate CLIMATES against the same baseline forecasting models that we used in Table 1, as all models there form the individual building blocks inside our CLIMATES approach. Note that we choose these algorithms as baselines in order to establish the effectiveness of our clustering based meta-algorithm approach over individual baselines. Further, we note that as more sophisticated time-series forecasting methods are developed in the AI community, they can also be utilized as building blocks inside our meta-algorithm approach. To have a fair comparison between various models, we conduct hyper-parameter tuning using the grid search approach. For complete details of experimental settings, please refer to the project's GitHub.

Table 5 shows the CV achieved by CLIMATES and all our baselines on single-step forecasting tasks on all three datasets. According to these results, CLIMATES outperforms all baselines on all datasets, e.g., CLIMATES achieves a CV of 0.0989 on the RET dataset, whereas the next best performing baseline achieved a CV of 0.1002. This establishes the superior performance of CLIMATES in providing accurate forecasts for AET, RET, and NPP. Additionally, we observe that the mentioned heuristic strategies for assigning data points to the clusters (i.e., CLIMATES-II and CLIMATES-III) leads to similar results. Note that although the improvement of CLIMATES over baselines does not look significant from an ML perspective, we will show, in the next section, that this improvement over baselines could result in considerable cost savings in the real-world.

| | AET | RET | NPP |
|---|---|---|---|
| TBATS | 0.2414 | 0.1206 | 0.2856 |
| SARIMA | 0.2130 | 0.1029 | 0.2503 |
| LR | 0.2114 | 0.1014 | 0.2492 |
| RF | 0.2080 | 0.1022 | 0.2427 |
| XGBoost | 0.2099 | 0.1039 | 0.2445 |
| SVM | 0.2110 | 0.1041 | 0.2505 |
| SFM | 0.2080 | 0.1002 | 0.2428 |
| TCN | 0.2138 | 0.1012 | 0.2432 |
| CLIMATES-I | 0.2075 | **0.0989** | **0.2409** |
| CLIMATES-II | **0.2071** | 0.0990 | **0.2409** |
| CLIMATES-III | **0.2071** | 0.0990 | **0.2409** |

Table 5: CV of CLIMATES and various baselines.

Orthogonally, Table 5 shows that although neural network models outperform popular statistical models by a relatively large margin ($\sim$1.78%, on average), their performance is comparable to some strong classical ML models. This finding is consistent with prior research, as there is a growing body of work which questions the superiority of some recent neural networks over classical ML models. For example, this finding is consistent with results reported in prior work in the area of Recommendation Systems (Dacrema, Cremonesi, and Jannach 2019), which found that some recent neural network models are not actually superior to well-tuned classical ML models. As an analogous result in the time-series forecasting domain, our findings suggest that despite the easy availability of large-scale datasets in time-series forecasting (due to easy access to remote sensing data), deep learning does not always beat traditional ML models significantly.

## Real-World Usage of CLIMATES

This section explains three possible ways that CLIMATES could be employed to help smallholder farmers in the field.

### Application 1: Forecasting Level of Water Stress

CLIMATES can be used to assist farmers in getting to know the occurrence of water stress in their farms ahead of time. Past literature suggests that water stress in each farm can be estimated from RET and AET using Equation 2 (Allen et al. 1998). In this equation, $K_s$ denotes the water stress index (e.g., $K_s < 0.5$ indicates an alarming level of water stress) and $K_c$ refers to the crop coefficient, for which the suggested values are available at (Allen et al. 1998). Thus, CLIMATES can serve as the ML engine of a mobile app that can send early warnings to farmers based on the future value of $K_s$ computed from the forecasted AET and RET.

$$K_s = \frac{AET}{K_c \times RET} \tag{2}$$

Further, the output of CLIMATES can be used to generate a heatmap, similar to Fig. 1b, to represent the water stress index across a large region. In this figure, the background color shows the forecasted level of water stress (assuming $K_c = 1.2$) across Kenya on the first dekad of May 2019 and black circles represents particular farm locations. According to this heatmap, the farms in western Kenya are at low-risk of water stress (as $K_s > 0.5$) during that particular dekad.

### Application 2: Irrigation Scheduling

CLIMATES can be utilized as an AI assistant for irrigation scheduling as well. Irrigation scheduling methods aim to determine the timing of irrigation and the amount of water demand at different stages of the crop-growing life cycle. One common approach in this space is ET-based irrigation scheduling, which utilizes ET data to provide customized suggestions for each farm based on its irrigation system, crop type, etc. According to this approach, the amount of water demand can be estimated using Equation 3 (Kisekka et al. 2019). In this equation, $GI$ denotes the gross irrigation water requirement, $ET_c$ denotes the crop evapotranspiration ($ET_c = K_c \times RET$), $P_e$ refers to the effective precipitation that can be consumed by plants, and $E$ denotes the efficiency of irrigation system used in the target farm. Thus, providing farmers with information on the future value of GI (through forecasting RET) can help them estimate the amount of water needed for mitigating the water stress in their farms.

$$GI = \frac{ET_c - P_e}{E} = \frac{(K_c \times RET) - P_e}{E} \quad (3)$$

**Real-World Impact of CLIMATES versus Baselines.** We now compare potential real-world impact of CLIMATES against the best-performing baseline model in the context of irrigation scheduling. To this end, we translate the amount of improvement in predictive accuracy (of CLIMATES over the best performing baseline) to the corresponding difference between GI (i.e., required levels of irrigation) computed from the outputs of CLIMATES (i.e., $\text{GI}_{CLIMATES}$) and the best-performing baseline (i.e., $\text{GI}_{baseline}$). This difference in GI ($\text{GI}_{baseline}$-$\text{GI}_{CLIMATES}$) could be an indicator of the amount of water that could be saved as a result of employing CLIMATES, rather than the best-performing baseline. However, translating the difference in predictive performance (in terms of CV) to the amount of water saving requires several assumptions as the value of CV does not distinguish under-estimation from over-estimation. In addition, the parameters of Equation 3 depend on various characteristics of the farm, e.g., $K_c$ changes with the crop type and the stage of crop growth. For ease of exposition, we assume that $E = 0.60$ (which corresponds to the Surface irrigation system (Brouwer, Prins, and Heibloem 1989)) and $K_c = 1.2$ (which corresponds to mid-season maize cropping (Allen et al. 1998)) are used in the target region, and that both CLIMATES and the best performing baseline (i.e., SFM) overestimate RET on a given dekad in that region. In this situation, according to Equation 4, the improvement of 0.0013 by CLIMATES over SFM in terms of CV in the RET prediction task (from Table 5) can be translated into saving about 92 liters of water per hectare each day for a maize-cropped farm at the mid-season stage. As a result, *although the improvement of CLIMATES against baselines looks small numerically, this improvement can result in considerable water saving when it comes to employing CLIMATES for scheduling irrigation within the crop growing season in the real-world.*

$$GI_{baseline} - GI_{CLIMATES} = \frac{K_c}{E}(RET_{baseline} - RET_{CLIMATES}) \quad (4)$$

### Application 3: Monitoring Crop Growth

CLIMATES can also be employed to quantitatively monitor plant growth. For example, NPP values forecasted by CLIMATES can be used to proactively identify some real-world stressors influencing plant growth such as nutrition shortage. In particular, CLIMATES can produce customized early warnings based on the amount of gap between the forecasted NPP and the NPP of the plant under non-stressed conditions.

## Challenges in Implementation

There are several challenges that need to be taken into account when planning for deployment in this domain. First, many African smallholder farmers live in rural areas with limited access to the Internet, and CLIMATES is an ML-based model in need of frequent updates. In fact, due to its computational needs, CLIMATES needs to be updated on a server with GPU. Therefore, access to the most recent information requires establishing a connection with a server via the Internet, and consequently, cannot be done offline. To address this challenge, we plan to add a feature to the app for automatically sending frequent updates to the registered farmers via text messages (SMS) so that they can stay updated even in case of Internet connection issues. The second challenge is related to the farmers' concerns about the privacy of their data. In fact, many farmers may not be willing to share some data such as farm size and crop type, as this information along with their estimated crop productivity could be used to derive their income, which is personal information to many people. We believe that PlantVillage which has already established trust with numerous African farmers could help mitigate this issue and encourage farmers' participation. We also plan to incorporate Federated Learning (Li et al. 2019) to enhance the protection of farmers' privacy.

## Conclusion

This paper proposes CLIMATES, an ML-based meta algorithm for forecasting three important crop-productivity related variables (AET, RET, and NPP) in smallholder farms across Africa. Leveraging structural insights about these variables, it attempted to combine the power of several popular time-series forecasting techniques to produce more accurate forecasts in the face of significant variability, mainly stemming from the geographic and climatic diversity of different African countries. The experimental results show that CLIMATES outperforms several strong baselines, including VRNNs which introduce latent variables to model variability in time-series data. CLIMATES is currently being evaluated by PlantVillage for potential future deployment as an early warning system for smallholder farmers in East Africa.

## Acknowledgments

This work was in part supported by NSF awards #1940076 and #1909702.

## References

Allen, R. G.; Pereira, L. S.; Raes, D.; and Smith, M. 1998. *Crop evapotranspiration : guidelines for computing crop water requirements*. FAO.

Alves, W.; Rolim, G.; and Aparecido, L. E. 2017. Reference Evapotranspiration Forecasting by Artificial Neural Network Models. *Engenharia Agrícola*, 37: 1116–1125.

Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Bowman, S. R.; Vilnis, L.; Vinyals, O.; Dai, A.; Jozefowicz, R.; and Bengio, S. 2016. Generating Sentences from a Continuous Space. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, 10–21.

Breiman, L. 2001. Random Forests. *Machine learning*, 45(1): 5–32.

Brouwer, C.; Prins, K.; and Heibloem, M. 1989. *Irrigation Water Management: Irrigation Scheduling*. FAO.

Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chien, J.-T.; and Kuo, K.-T. 2017. Variational Recurrent Neural Networks for Speech Separation. In *INTERSPEECH*, 1193–1197.

Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A. C.; and Bengio, Y. 2015. A recurrent latent variable model for sequential data. *Advances in neural information processing systems*, 28: 2980–2988.

Cortes, C.; and Vapnik, V. 1995. Support-Vector Networks. *Machine Learning*, 20(3): 273–297.

Dacrema, M. F.; Cremonesi, P.; and Jannach, D. 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*, 101–109.

Del Grosso, S.; Parton, W.; Stohlgren, T.; Zheng, D.; Bachelet, D.; Prince, S.; Hibbard, K.; and Olson, R. 2008. Global potential net primary production predicted from vegetation class, precipitation, and temperature. *Ecology*, 89(8): 2117–2126.

FAO. 2018. *WaPOR Database Methodology: Level 2. Remote Sensing for Water Productivity Technical Report: Methodology Series*. Rome: FAO.

Feng, K.; and Tian, J. 2021. Forecasting reference evapotranspiration using data mining and limited climatic data. *European Journal of Remote Sensing*, 54(sup2): 363–371.

Gollin, D. 2014. *Smallholder agriculture in Africa: An overview and implications for policy*. International Institute for Environment and Development (IIED).

Harvey, C.; Rakotobe, Z. L.; Rao, N. S.; Dave, R.; Razafimahatratra, H.; Rabarijohn, R.; Rajaofara, H.; and MacKinnon, J. L. 2014. Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1639): 20130089.

He, J.; Spokoyny, D.; Neubig, G.; and Berg-Kirkpatrick, T. 2019. Lagging Inference Networks and Posterior Collapse in Variational Autoencoders. In *International Conference on Learning Representations (ICLR)*.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Hyndman, R. J.; and Khandakar, Y. 2008. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3): 1–22.

Izadifar, Z. 2010. *Modeling and Analysis of Actual Evapotranspiration using Data Driven and Wavelet Techniques*. Master's thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Jia, X.; Wang, M.; Khandelwal, A.; Karpatne, A.; and Kumar, V. 2019. Recurrent generative networks for multi-resolution satellite data: an application in cropland monitoring. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2628–2634.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations (ICLR)*.

Kisekka, I.; Migliaccio, K. W.; Dukes, M. D.; Schaffer, B.; Crane, J. H.; Bayabil, H. K.; and Guzman, S. M. 2019. Evapotranspiration-Based Irrigation Scheduling for Agriculture. https://edis.ifas.ufl.edu/pdffiles/AE/AE45700.pdf. [Online; accessed 12-January-2021].

Landeras, G.; Ortiz-Barredo, A.; and López, J. J. 2009. Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. *Journal of irrigation and drainage engineering*, 135(3): 323–334.

Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; and He, B. 2019. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *arXiv preprint arXiv:1907.09693*.

Livera, A. M. D.; Hyndman, R. J.; and Snyder, R. D. 2011. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496): 1513–1527.

Rezende, D. J.; Mohamed, S.; and Wierstra, D. 2014. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *Proceedings of the 31st International Conference on International Conference on Machine Learning (ICML)*, 1278–1286.

Roelofsen, P. 2018. *Time series clustering*. Master's thesis, Vrije Universiteit Amsterdam, Netherlands.

Salami, A.; Kamara, A. B.; and Brixiova, Z. 2010. *Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities*. Tunis, Tunisia: Working Papers Series N° 105 African Development Bank.

Shimeles, A.; Verdier-Chouchane, A.; and Boly, A. 2018. Introduction: understanding the challenges of the agricultural sector in Sub-Saharan Africa. In *Building a Resilient and Sustainable Agriculture in Sub-Saharan Africa*, 1–12. Springer.

Sun, J.; and Du, W. 2017. Effects of precipitation and temperature on net primary productivity and precipitation use efficiency across China's grasslands. *GIScience & Remote Sensing*, 54(6): 881–897.

Zhang, L.; Aggarwal, C.; and Qi, G.-J. 2017. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2141–2149.

Zhang, S.; Zhang, R.; Liu, T.; Song, X.; and Adams, M. 2017. Empirical and model-based estimates of spatial and temporal variations in net primary productivity in semi-arid grasslands of Northern China. *PLoS ONE*, 12(11): e0187678.