Global custom-tailored machine learning of soil water content for locale specific irrigation modeling with high accuracy

Aadith Moorthy

MSC 697, California Institute of Technology, Pasadena, CA 91126, USA
Email: amoorthy@caltech.edu

Abstract**—***A novel approach to irrigation modeling is presented: the locale specific machine learning of soil moisture data. The merits of this new patent pending technique are clear when compared to existing methods, such as the AquaCrop program created by the Food and Agricultural Organization (FAO). From a case study on the comparative performance of AquaCrop and machine learning in the extrapolative modeling of soil moisture, AquaCrop performed with a mean squared error of 0.00165 whereas the machine learning received 0.00013, an order of magnitude lower. In addition, a novel algorithm, the ConserWater™ algorithm, has been created for the purpose of machine learning soil moisture with accuracy and efficiency. The performance of the algorithm is very superior when compared to other popular machine learning techniques, as applied to soil moisture. Finally, to allow this technology to reach agriculturalists at the grassroots level, the entire world has been machine learned and the resultant models have been encapsulated into a lightweight easy-to-use smartphone application.*

Keywords— ConserWater™, irrigation, machine learning, soil moisture.

I. INTRODUCTION

The Earth has 1.4 billion cubic kilometers of water, but at any given time only about 200,000 cubic kilometers is fresh water that is accessible for human use [1]. In addition to the inherent paucity of total water supplies in relation to the escalating human population, anthropogenic traumata have already taken a significant hold. This is especially true in agricultural regions, as agriculture is the largest drain of fresh water [2].

A recent UN report has shown that small farms, preeminent in the developing world, provide the majority of food worldwide, and employ 3 billion people [3]. In addition, most of the people living below the dollar-a-day poverty line are a part of this group [3]. As they cannot afford sprinklers or drip irrigation for their fields, they resort to the age-old inefficient practice of surface irrigation: they flood their fields with significant water and wait for it to drain into their fields. As a result, not surprisingly, nearly 2/3 of global fresh water is used for irrigation [4]. It is by uplifting small farmers in developing countries that we can achieve another agricultural revolution and ameliorate global water consumption while optimizing plant growth.

It is envisaged that with good management and knowledge of the optimum quantity of water needed for irrigation, the efficiency of even the surface irrigation method can be driven up to as high as 90% [5]. To this end, several companies sell field-wide soil moisture sensor arrays to provide real-time data on crops water needs [6]. This approach can be efficient and it is claimed that it can facilitate a 25% decrease in water consumption during drought [6]. However, it can also be prohibitively expensive, except for wealthy farmers. Each moisture sensor can cost about $100 dollars, so implanting an array, along with operational services, can cost several thousand dollars [7]. This is clearly beyond the budgets of small farmers.

As an alternative, more cost effective, approach, the irrigation science community has been researching the usage of mathematical models for irrigation management for the past decades. One of the most notable of these models is AquaCrop, a crop water model developed by the Land and Water Division of the Food and Agricultural Organization (FAO) [8]. For the purpose of predicting irrigation water requirements, the central postulation used by AquaCrop is the soil water balance [9]:

$$D_{r,i} = D_{r,i-1} - (P - RO)_i - I_i - CR_i + ET_{c,i} + DP_i \quad (1)$$

where $D_{r,i}$ represents the root zone water depletion on day i, $D_{r,i-1}$ the root zone water depletion on day $i-1$, P_i the precipitation on day i, RO_i runoff from the soil surface on day i, I_i net irrigation depth on day i, CR_i capillary rise from the groundwater on day i, $ET_{c,i}$ crop evapotranspiration on day i and DP_i water loss of the root zone by deep percolation on day i.

From (1), it is clear that the inputs of water into the soil are rainfall, irrigation and capillary rise, whereas the outputs are deep percolation, runoff and evapotranspiration (ET). Of the output mechanisms, ET is the most important, as the others are only significant when there is heavy rainfall or heavy irrigation [9]. One drawback of this equation is that it implicitly assumes a purely vertical model of water movement, whereas there may be other factors that distort this simple model.

However, more importantly, it turns out that ET is exceedingly difficult to calculate with accuracy. For reference ET, which simulates the evapotranspiration that occurs from a hypothetical .12m grass-like crop, under conditions of ample water, the FAO recommends and endorses the FAO 56 Penman-Monteith Equation [9]. It is well known that this equation can produce unreliable results when climate records are sparse or unavailable [10]. This is becoming less of an issue in the 21st Century as numerous Autonomous Weather Stations (AWS) are being placed around the world, even in developing countries, like China, which has over 4000 weather stations, sufficient to cover its entire geographical realm [11]. But some crucial parameters are not well tabulated for use in the equation: for example, solar radiation and wind speed at a height of 2 m above the ground are required by the model, but these quantities are seldom measured and are usually estimated in a crass manner [12, 13]. Additionally, ground heat flux is a parameter needed by the model that is similarly very difficult to accurately measure [14].

A common issue with any general equation approach is the presence of confounding factors, and geographical variability, which has been well documented in the literature [15]. It has recently been suggested that even the Penman-Monteith equation is not immune to problems upon closer inspection [16]. There are alternative general equations, such as the Hargreaves equation, but the problems remain, and detract from the usage of a general equation globally to predict reliable ET [15].

Accordingly, there are several other methods for ET determination, such as lysimetry, pan evaporation and empirical-based statistical modeling [9]. The usage of lysimeters to determine ET is the most direct method, but it is prohibitively expensive to install and use, and is hence not widespread [9]. Pan evaporation is easy to implement but is not favored, because environmental factors differ between pan evaporation and soil with crops growing [9]. It can only give a qualitative idea of ET.

Statistical modeling has been a relatively new approach facilitated by the advent of computers that can vastly simplify the modeling process. Since ET tends to be a nonlinear behavior in several variables, Artificial Neural Networks (ANNs) are commonly used [17]. The details of how ANNs work, especially in an irrigation context, is well documented in the literature [18]. The models produced by ANNs can be made to be very accurate in predictive power. However, the necessary drawback is that they can become computationally intensive as more neurons are supplied to increase their modeling power. Despite advances in computer technology, there is currently no way for ANNs with several hidden neurons to be efficiently deployed and executed on personal computers and smartphones, as it is well known they require large amounts of parallel processing on distributed computing systems for efficacy [17]. On the other hand, if a small number of neurons are used, the model is bound to not be very accurate as ET is a nontrivial function of the input variables. Hence ANNs may not be useful for extensive modeling of every locale worldwide, which is required to allow irrigation management technology to reach every farmer.

A final roadblock to the statistical modeling of ET is where the training data comes from. There has to be a set of known ET values and corresponding weather data to train the models. However, as discussed before, it is often not feasible to measure ET directly. As a result, studies on training statistical models of ET usually utilize the Penman-Monteith equation to produce the training values [18]. Therefore, any potential problems from the usage of the equation, possibly from inadequate weather data, are carried over to the models. Indeed, some of these studies also state that these models are intended to be used in more

of a 'gap-filling' procedure to fix missing data, rather than be an efficient calculator of evapotranspiration at all times [17, 19].

In this paper, a radical new approach is taken to irrigation modeling, by completely eschewing the soil water balance as the crux of the model. Without this balance at the center, there is no need to focus most efforts on acquiring ET, which, as described so far, is very hard to calculate accurately. Instead, the new approach is to directly machine learn soil moisture data. The methods of generating soil moisture values from remote sensing are well known, and can generate values to an accuracy of 0.01 m^3/m^3, which is sufficient for most applications [20, 21]. Furthermore, remote sensing enables such reasonably accurate values of soil moisture to be determined for any place on earth. Therefore, there is no need for one general model to describe everything: separate models can be generated for each location, so local variables are controlled by the models themselves.

The machine learning of soil moisture is a very difficult task, as it is affected by all the variables in the soil water balance from (1). In addition, there may be hidden variables not accounted for or well described by (1) that could potentially affect soil moisture. It is shown in this paper that several standard and advanced machine learning techniques, such as multiple linear regression and ensemble learning do not fair well in describing the data. The data is technically a time series, but dedicated time series techniques, such as the popular AutoRegressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) also do not work because the stationarity assumption is not satisfied. This assumption requires that the dataset have statistical properties, such as mean and standard deviation, that remain constant over time [22]. In SARIMA, seasonal variation is allowed, as long as it predictably repeats itself in the same manner [22]. However, this is not the case for any data that is controlled by precise climatic factors, as is any calculation of irrigational interest. There is also no way to accurately discount these climatic factors to create processed data that is stationary. Some prior irrigation modeling related work has tried to consider all non-seasonal variation to be random, but this may be an inadequate approach [23]. Climatic variation is not purely random, as it is affected by several localized variables in a given region. This is especially important for daily simulations and in the modeling of peaks in data resulting from precipitation, as is done here. Therefore, such an approach is bound to give inaccurate results when used in a real application.

To combat these difficulties in the machine learning of soil moisture, a novel algorithm, the ConserWater™ algorithm, has been developed. This algorithm not only shows high accuracy in modeling, but also good computational efficiency, to ensure the models are learned in an expedient manner. In this paper, the ConserWater algorithm is compared to several popular machine learning techniques, and also the FAO endorsed AquaCrop application, to attest to its efficacy.

In fact, the algorithm is so efficient without compromises in accuracy that it can be used to develop separate models for every location on earth. The widespread geospatial machine learning has already been carried out, and thousands of models have successfully been generated for several regions of the world. All of these models have been encapsulated in a lightweight easy-to-use smartphone application, to enable this technology to reach farmers and agricultural engineers at the grassroots level [24]. The particular details of the data mining, geospatial machine learning, and smartphone application are also described.

II. MATERIAL AND METHOD

2.1 Geospatial Machine Learning and Data Mining

Using the Global Surface Summary of the Day (GSOD) dataset produced by the National Oceanic and Atmospheric Administration (NOAA) of the US, the locations of over 30,000 reliable weather stations worldwide were received [11]. Soil moisture data was acquired from the National Snow and Ice Data Center (NSIDC) [23]. Surface soil moisture was data mined for each of the vicinities of the 30,000 stations, from datasets available from NSIDC, in three-hour increments, for the course of over a year from 3/31/15 to 4/21/16. Hence, a time series collection of soil moisture data was created for each station.

The data were then shifted and adjusted to account for time zonal differences. Next, the three-hourly data were averaged to obtain daily data, to remove intraday variation resulting from hourly changes in temperature and sunlight.

The daily data, which consisted of approximately 400 data points for each station, was later machine learned using the ConserWater™ algorithm. The physical machine learning was performed using Scikit-learn, a well known set of python tools for machine learning and statistics. 10% of points from the almost 400 data points for each station was randomly selected for cross validation testing (10-fold cross validation) and was not used as part of the training set. The inputs to the machine learning consisted of location specific weather data collected from the stations themselves. Some examples of the input variables used include temperature, pressure, and rainfall. This data was acquired using the World Weather Online service.

Using the cross validation set of 10% of the data points, mean squared errors (MSE) were calculated as follows:

$$MSE = \frac{\sum_{i=1}^{N}(y-\hat{y})^2}{N} \tag{2}$$

where N represents the number of data points, y is the real data value and $\hat{y}$ is the predicted value. The MSE quantifies how well the model fits the data, and is commonly used in the machine learning literature.

In general, it was ensured that the MSE in soil moisture for each model remained below 10^{-4}, to ensure the quality of the models. If any were found to not reach this threshold, those models were retrained until they converged below this level. In particular, 10^{-4} was chosen because it was close to the random error in the soil moisture data. Any features greater than 10^{-2} are likely to be real features of the soil moisture, so the MSE tolerance was set to $(10^{-2})^2 = 10^{-4}$.

Accounting for various crop factors and rooting depths during all stages of crop growth, which have been well documented, the water needs of these plants can be accurately predicted, in order to ensure optimal growth with just enough water [25].

The 30,000 models for the various different weather stations and their vicinities were encoded into a light-weight and easy to use android application, ConserWater™, for facile viewing [24].

2.2 Comparison of various Machine Learning algorithms

In order to test and compare the efficacy ConserWater™ algorithm and process, as is presented in this paper, several additional well known algorithms and methods were also tried: Least Squares Multiple Linear Regression (LSMLR), Sequential Feature Selection (SFS), LASSO Regression, Ridge Regression, Elastic Net Regression, Ensemble Learning and Neural Networks. They were all learned with the same input data. MATLAB was used to produce these comparative models.

LSMLR was the simplest algorithm tried and involves the minimization of residual squares, to find optimal coefficients:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad s.t. \quad \min_{\beta_0,\ldots,\beta_n} \sum_{i=0}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

where $\beta_0, \ldots, \beta_n$ are the fitted coefficients and $x_1, \ldots, x_n$ represent the regressors of y. N is the total number of data points.

The SFS algorithm is similar to LSMLR, except that an additional underlying procedure removes input variables that do not affect the output with statistical significance, effectively reducing the computational complexity of the machine learning while keeping the quality of the models fairly similar. For this study, an F-test was used to determine significance.

LASSO is an abbreviation for "least absolute shrinkage and selection operator". It is an algorithm that also does variable selection, similar to SFS. But it additionally invokes regularization. Regularization is a statistical technique that specifies additional information to avoid overfitting of data. Mathematically, LASSO is described as follows:

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 \quad s.t. \quad \sum_{j=1}^{n}|\beta_j| \leq t \tag{4}$$

where t is a parameter that quantifies the regularization employed in the model, and x_i^T represents the transpose of x_i.

Ridge Regression is similar to LASSO, as it also employs regularization to combat overfitting. However, it does not allow any coefficients to be set to zero, whereas the LASSO can. Ridge Regression is mathematically described:

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 \;\; s.t. \;\; \sum_{j=1}^{n}\beta_j{}^2 \leq t \tag{5}$$

The crucial difference between the LASSO and Ridge Regression develops from the difference in the inequality.

The elastic net, developed in 2005, is a hybrid algorithm encompassing both Ridge Regression and the LASSO. It has a built-in parameter, α, which quantifies the extent to which the algorithm resembles the LASSO or Ridge Regression. For example, if $\alpha = 1$, the algorithm resembles the LASSO, and for $\alpha = 0$, Ridge Regression. In this study $\alpha = .5$ was set.

Ensemble learning, one of the most promising techniques, is an umbrella term for techniques that use an array of learners, which are combined to give the final result. The advantage of this procedure is that the likelihood of a poorly trained model is reduced when several learners are combined. The most commonly used ensemble techniques are bagging and boosting.

Bagging is an acronym for 'bootstrap aggregating', and it involves a statistical technique called bootstrapping to obtain several randomly generated datasets from the originally supplied input regressor dataset. Then, separate models are made for each of these new datasets. Each of these models is called a 'learner' and their results are combined with equal weighting to give a prediction. In this study, 100 learners were used.

Boosting is very similar to bagging, except in the creation of new datasets from the original input regressor data. The new datasets are strategically created such that up to half of the data is well described by the previously developed learners, and the rest is not. Therefore, this method allows the model to drastically improve as the number of learners increases. Similar to the bagging, 100 learners were also used for the boosting in this study.

The machine learning procedure that has probably received the most media and scholarly attention is that of neural networks. They are commonly described as modeled after the human central nervous system. Several publications have detailed how neural networks work and how they have been used to model quantities of irrigational interest [17, 18]. In particular, a general regression neural network with the Levenberg-Marquardt algorithm was chosen and the number of hidden neurons was set to 30 to obtain a trained model that was temporally efficient. 80% of the input data was used for training, with 10% validation and 10% testing. These are also similar to those employed for calculations of irrigational interest in prior work [17].

As mentioned before, machine learning techniques that are often used with time series data, such as the autoregressive integrated moving average (ARIMA), were not used because the data did not satisfy their assumptive requirements.

All of the aforementioned techniques are presented for data from Coimbatore, Tamil Nadu, India. The models were generated, trained and tested using soil moisture data from 3/31/15 to 4/21/16, which was acquired as mentioned before.

2.3 Comparison with Aquacrop

For a comparison of the ConserWater™ algorithm with existing technologies for irrigation management, case studies were also conducted with AquaCrop, the program created by the FAO. In this paper, a particular case study of Amaravathi, Maharashtra, India is presented. Identical weather data was inputted to both the ConserWater™ algorithm and AquaCrop. For the trends in ET needed by AquaCrop over the course of the study period, the Penman-Monteith Equation was used:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{6}$$

The variables shown in (6) are as defined in [26]. For the sake of simplicity, it was assumed that there were no crops substantially different from the hypothetical reference, which was justified by satellite imagery of the area, and the area was also not known to be under climatic stress, so the reference ET was the real ET. Also, the soil type at Amaravathi was determined from the NSIDC datasets to be silt clay loam. Finally, again for simplicity, no irrigation procedure was entered, so that the predicted results could be compared to real data from the NSIDC dataset. Using an initial value that was itself from the dataset, the AquaCrop model was propagated forward and the trend in surface soil moisture from the model was recorded. The ConserWater™ model was also propagated using the same initial values and data. Finally, to provide a holistic

view of ConserWater™ in relation to AquaCrop, the MSE found for 10 more locations are also presented, following the same methods.

III. RESULTS AND DISCUSSION

3.1 Comparison of various Machine Learning algorithms

The predictions of the various developed machine learning models for Coimbatore are presented in Fig.1, with regressors supplied to the models just as they were trained (no future extrapolative behavior). Only a small segment of the almost 400 days is shown in order to emphasize how well the data is being described by each of the models.

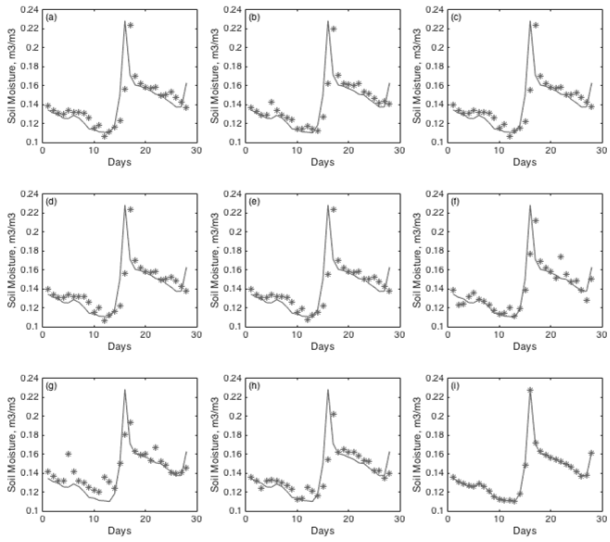

FIG.1. PREDICTIONS OF VARIOUS SOIL MOISTURE MACHINE LEARNING MODELS: (A) MLR, (B) SFS, (C) THE LASSO, (D) RIDGE REGRESSION, (E) ELASTIC NET, (F) BAGGED ENSEMBLE, (G) BOOSTED ENSEMBLE, (H) NEURAL NETWORK AND (I) CONSERWATER™. THE SOLID LINE IS THE REAL DATA, WHILE THE DOTTED LINE CONSISTS OF THE PREDICTED VALUES.

On a perfunctory glance, it appears as though all the models do a satisfactory job. Most of the fits look similar, especially the LSMLR, SFS, LASSO, Ridge and elastic net. The ensemble learning methods give a marginally worse prediction, as does the neural network. The ConserWater™ model gives by far the best fit, but all the models seem to be adequate for any purpose of soil moisture prediction, because their residuals are generally low and R^2 values generally in the .95-.99 range. The MSE for these models is shown in Fig.2, which quantitatively confirms how each of these models describes the data.

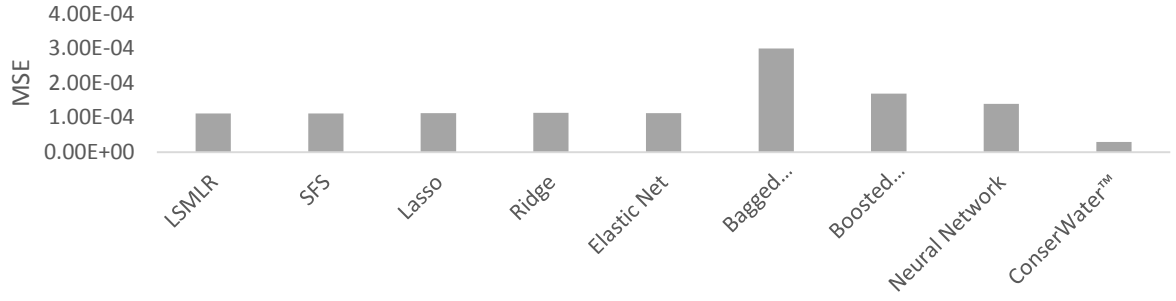

FIG.2. MEAN SQUARED ERRORS (MSE) FOR THE VARIOUS MACHINE LEARNING MODELS AND DATA, AS SEEN IN FIG.1.

It may seem initially that the ConserWater™ model is "overfitting" the data. Overfitting is the common machine learning error where the model seems to unnaturally fit the data, to the point of capturing variations due to noise. With this reasoning,

one may believe that the other models are better, as they do not capture the noise. However, this reasoning is unfounded for two major reasons: (1) the ConserWater™ model is not overfitting and (2) the others models have a common issue.

In the case of soil moisture data, there is an inherent daily change caused by a variety of mechanisms such as evaporation and rainfall. From this view, the ConserWater™ model effectively follows such trends with precision. Whereas other models show upward trends in soil moisture even when there is no water input into the soil, the ConserWater™ model avoids this behavior. Furthermore, the percent error in the soil moisture data for Coimbatore, as quoted in the dataset, is just 10%, so most features in the real data are features from the moisture, and not instrumental error. In fact, an observation of the three hour dataset, before it was averaged, reveals the daily fluctuations in soil moisture and rate of change. These fluctuations are ~0.01 m^3/m^3, so they suggest that signal noise must be smaller than 0.01 m^3/m^3, because they would otherwise be invisible.

Similarly, the egregious issue in the other models is that they do not accurately describe peaks. Peaks can be caused by rainfall, as in this data, or even by irrigation, which is of practical relevance for this study. Models are of no use if they cannot describe when the soil moisture will reach field capacity during irrigation, to avoid wasteful consumption of water. Upon close inspection, in Fig.1, it can be seen that the predicted peak values of soil moisture in all models other than the ConserWater™ model tend to occur a day after the real peak. This indicates the models have not actually learned how rain peaks work, but are just returning the previous value that was given to them in the input. This may be the reason why there is no prior literature on machine learning soil moisture itself: there are complexities surrounding the peaks. This issue is summarized in Fig.3, where the regressors provided to the models were altered to use extrapolation to calculate soil moisture in a self-iterative fashion. This is how these models would be used in the real world for predictions days after irrigation.

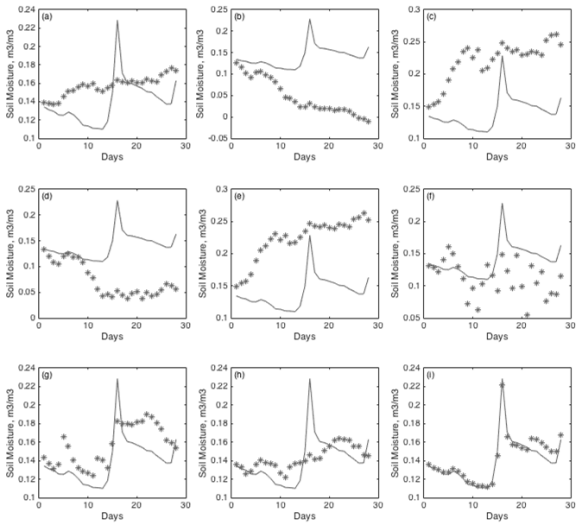

FIG.3. EXTRAPOLATED PREDICTIONS OF VARIOUS MACHINE LEARNING MODELS, WHEN ONLY PROVIDED AN INITIAL SOIL MOISTURE VALUE: (A) MLR, (B) SFS, (C) THE LASSO, (D) RIDGE REGRESSION, (E) ELASTIC NET, (F) BAGGED ENSEMBLE, (G) BOOSTED ENSEMBLE, (H) NEURAL NETWORK AND (I) CONSERWATER™. THE SOLID LINE IS THE REAL DATA, WHILE THE DOTTED LINE CONSISTS OF THE PREDICTED VALUES.

Fig.3 reveals the fallacy in utilizing well known methods for machine learning moisture. Models that seemed to fit the data well now no longer do so, and several diverge. Now the LSMLR, SFS, LASSO, Ridge and elastic net all seem to err on too low or too high values and have lost the day 16 rain peak. The bagged learning model produces a random scatter. The boosted learning, mentioned earlier to be one of the promising algorithms, has mixed success. It understates the rain peak on day 16 while overstating the small rain "bulge" from day 6. The neural network model seems to closely follow all parts of the data except for the rain peak. Finally, ConserWater™ performs best. It fits the data less than in Fig.1, but is of much higher value as it follows all the trends, plus the rain peak. Therefore, the ConserWater™ algorithm is a suitable algorithm for accurate prediction of moisture. These observations can be quantitatively seen in Fig.4, which shows model MSEs.

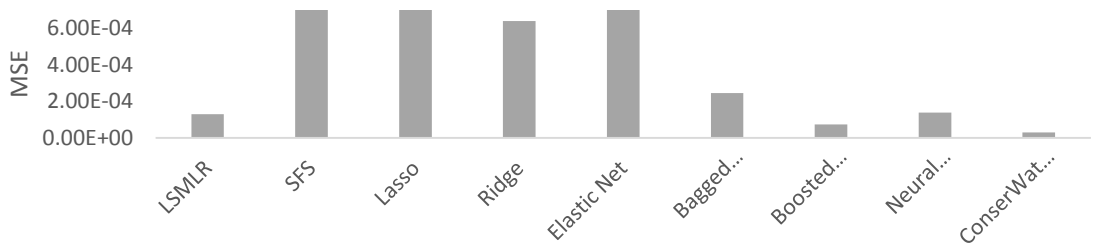

FIG.4. MEAN SQUARED ERRORS (MSE) FOR THE VARIOUS MACHINE LEARNING MODELS AND DATA, AS SEEN IN FIG.3.

3.2 Comparison with AquaCrop

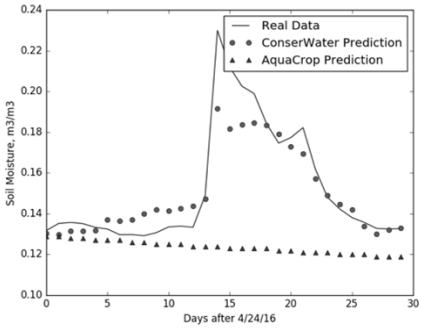

FIG.5. EXTRAPOLATION OF CONSERWATER™ MODEL AND OUTPUT OF AQUACROP FOR SURFACE SOIL MOISTURE, COMPARED TO REAL DATA.

AquaCrop uses a basic water balance, where rainfall and ET are primary factors. The rainfall at Amravati on Day 15 in Fig.3 was just 2 mm, but was sufficient to produce a rain peak. Clearly ET values around 15 mm/day exceed 2 mm, and the water balance yielded incorrectly that soil moisture would drop. Note that even if the ET were incorrectly calculated due to erroneous weather data, there is still no approach for the AquaCrop model to get close to the same rain peak as the real data, since 2 mm added to the soil moisture cannot result in a ~.1 increase in soil moisture for any reasonable depth of soil. This only reveals the robustness of the ConserWater algorithm with erroneous data, as compared to AquaCrop. Therefore, Fig.3 clearly reveals the inadequacies of AquaCrop: not only are ET based on general equations unreliable, but the water balance is also not sufficiently accurate to predict how the soil moisture rises with just a little rain, especially with potential inaccurate data. Hence AquaCrop is lacking for location-specific global application, further supported by results for more locations:

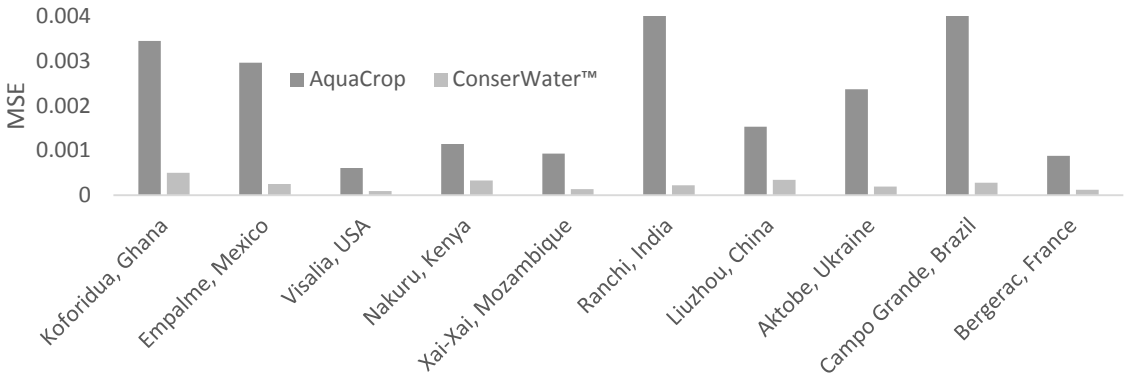

FIG.6. MSE OF AQUACROP AND CONSERWATER™ FOR VARIOUS LOCATIONS.

In context, the MSE values from Fig.6 point out that soil moisture values from AquaCrop are typically within $\sim\sqrt{0.0025} = \sim 0.050$ of the right value, whereas ConserWater has an error margin around $\sim\sqrt{0.00025} = \sim 0.016$. This may not seem large now, but this difference can equate to tens of thousands of gallons of irrigation water for just a single acre.

3.3 The Smartphone Application

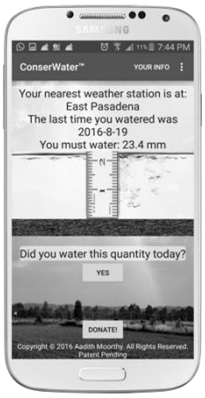

FIG.7. SCREENSHOT FROM THE CONSERWATER™ APPLICATION, DISPLAYING THE REQUIRED IRRIGATION FOR PASADENA ON A GIVEN DAY.

In order to make this technology accessible to farmers and agricultural engineers, it has been developed into an Android/iOS application. With user inputs similar to AquaCrop, but more user-friendly, this application gets the user's location from the location services of a smartphone. Using the location, the application selects the most representative station model. Some factors taken into account in selecting the model to be used are proximity, climate compatibility and soil similarity.

Additionally, the user is allowed to select the initial value for the soil moisture in the prediction. The user may enter the last date he or she irrigated the field extensively, so the water content is assumed to be at the field capacity on that date, and the model is subsequently evolved. Otherwise if the user has not watered his or her field recently, the soil moisture initial value is taken to be the most recent value for the location from the NSIDC datasets. Finally, the selected model is propagated forward in time using the initial values and recent historical weather data to predict the current levels of soil moisture. The current levels can then be used with crop factors to determine water requirements. If utilized, this water quantity will bring the soil moisture back to an optimum. A screenshot from a prediction of the application is displayed in Fig.7.

An important aspect is that it is lightweight – even phones with low specifications can run it. Furthermore, the models are computationally efficient, during training and prediction, enabling the machine learning of the world, and quick responses to usage. A recent trial of this application with Indian and Ghanaian farmers has also resulted in positive appraisals.

IV. CONCLUSION

A monumental algorithm capable of efficiently machine learning soil moisture data to high accuracy has been presented. Compared to several popular existing techniques, it clearly has the upper hand in terms of accuracy. Additionally, it can provide a much more precise model of localized effects than programs such as AquaCrop, that rely on a simple water balance and evapotranspiration calculation as intermediate steps. This level of preciseness can translate to savings of tens of thousands of gallons of water for just an acre of land. Finally, this technology has been encapsulated into an easy-to-use lightweight smartphone application. These are necessities for the widespread proliferation of irrigation management technology around the world, especially to small farmers who cannot afford soil moisture sensors or other irrigation management technology. More information on using this technology, including download links, can be found on www.conserwater.com.

REFERENCES

[1] Boberg J (2006) One world, One well: How Populations Can Grow on a Finite Water Supply. In: Rand Corporation. http://www.rand.org/pubs/periodicals/rand-review/issues/spring2006/water.html. Accessed 3 Aug 2016

[2] Postel SL, Daily GC, Ehrlich PR (1996) Human appropriation of renewable fresh water. Science 271:785–788.

[3] IFAD, UNEP (2013) Smallholders, food security, and the environment.

[4] Wolfe B, Berger B, Filiberto D, Newton M, Pimentel D, Karabinakis E, Clark S, Poon E, Abbett E, Nandagopal S (2004) Water resources: Agricultural and environmental issues. BioScience 54:909–918. doi: 10.1641/0006-3568(2004)054[0909WRAAEI]2.0.CO;2

[5] Henkel M (2015) 21st century homestead: Sustainable agriculture III: Agricultural practices. Lulu.com

[6] Lufkin B (2015) Soil sensors can cut farms' water use by a quarter during drought. In: Gizmodo. Accessed 3 Aug 2016

[7] Sanden B (2005) Making Sense of Soil Moisture Checking and Sensors. University of California, Bakersfield, CA, USA

[8] Steduto P, Hsiao T, Fereres E, Raes D (2012) Crop yield response to water. Food and Agriculture Organization of the UN, Rome, Italy

[9] Allen R, Pereira L, Raes D, Smith M (1998) Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations, Rome, Italy

[10] Stöckle CO, Kjelgaard J, Bellocchi G (2004) Evaluation of estimated weather data for calculating Penman-Monteith reference crop evapotranspiration. Irrigation Science 23:39–46. doi: 10.1007/s00271-004-0091-0

[11] NOAA (2016) Global surface summary of the day - GSOD. NOAA data catalog

[12] Llasat M.(1998)Data error effects on net radiation and evapotranspiration estimation. Agricultural and Forest Meteorology 91:209–221.

[13] Jensen DT, Hargreaves GH, Temesgen B, Allen RG (1997) Computation of ETo under Nonideal conditions. Journal of Irrigation and Drainage Engineering 123:394–400. doi: 10.1061/(asce)0733-9437(1997)123:5(394)

[14] Wang J, Bras R. (1999) Ground heat flux estimated from surface soil temperature. Journal of Hydrology 216:214–226.

[15] Temesgen B, Eching S, Davidoff B, Frame K (2005) Comparison of some reference Evapotranspiration equations for California. Journal of Irrigation and Drainage Engineering 131:73–84. doi: 10.1061/(asce)0733-9437(2005)131:1(73)

[16] Schymanski SJ, Or D (2016) The failure of the Penman-Monteith equation in explaining leaf transpiration. EGU General Assembly Conference Abstracts 18:12635.

[17] Martí P, Gasque M (2010) Reference evapotranspiration estimation without local climatic data. Irrigation Science 29:479–495.

[18] Kim S, Kim HS (2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. Journal of Hydrology 351:299–317. doi: 10.1016/j.jhydrol.2007.12.014

[19] Abudu S, Bawazir AS, King JP (2010) Infilling missing daily Evapotranspiration data using neural networks. Journal of Irrigation and Drainage Engineering 136:317–325. doi: 10.1061/(asce)ir.1943-4774.0000197

[20] Nichols S (2011) Review and evaluation of remote sensing methods for soil-moisture estimation. Journal of Photonics for Energy 028001. doi: 10.1117/1.3534910

[21] Reichle, R., G. De Lannoy, R. D. Koster, W. T. Crow, and J. S. Kimball. 2016. *SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 2*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Archive Center.

[22] Brillinger D (1975) Time series: Data analysis and theory. Holt, Rinehart & Winston

[23] Psilovikos A, Elhag M (2013) Forecasting of remotely sensed daily Evapotranspiration data over Nile Delta Region, Egypt. Water Resources Management 27:4115–4130. doi: 10.1007/s11269-013-0368-2

[24] Moorthy A (2016) ConserWater. In: ConserWater. https://www.conserwater.com. Accessed 20 September 2016

[25] Guerra E, Ventura F (2016) Crop coefficients: A literature review. Journal of Irrigation and Drainage Engineering 142:06015006.

[26] Zotarelli L, Dukes M, Romero C, Migliaccio K, Morgan K (2015) Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). University of Florida Institute of Food and Agricultural Sciences, Gainesville, Florida, USA