Utilizing a Long Short-Term Memory (LSTM) Machine Learning Algorithm to Create Soil Moisture Prediction Models and Improve Water Productivity in Southern California

Aurrel Bhatia

Personal Section

When serializing Earth's evolution with a clock, we can image that everything developed, evolved, and advanced within 23 hours and 59 seconds, and human existence comparatively makes up that last second. When I first learned this fact, I immediately thought about how in the grand scheme of things, human have had an extremely short time to impact our world. This does not however, mean that the impact has been small by any means. Human activity such as pollution, overpopulation and overconsumption of natural resources, and industrialization, can be directly linked to habitat destruction, deforestation, overall environmental degradation, and climate change.

Regarding utilizing resources, human activity exhibits two extremes of a spectrum (overconsumption and scarcity), demonstrated by how we consume the most basic substance: water. Generally, we perceive water as an infinite resource, one that can be used almost indiscriminately. The truth is far from that, since only 1% of all available water on Earth is actually usable for human activity. Even with this limited amount, due to our current technological, agricultural, and economic systems, many disparities in access to freshwater for drinking and daily activities exist, further exacerbating water scarcity, or the lack of sufficient freshwater resources to meet the demand in a region. In many instances, water scarcity is caused or propagated by developed and heavily industrialized nations participating in excessive consumption, inhibiting other nations from accessing basic necessities.

Currently, two-thirds of the global population experiences water scarcity. One-fifth of the world population currently lives in conditions of physical water scarcity, where there is not enough water to meet their demands, and one-quarter of the world's population experiences economic water scarcity, where their region has enough water to meet the necessary personal, agricultural, environmental, and industrial needs, but lack sustainable accessibility. Existing literature determined that water scarcity occurred more often in areas where irrigation systems had low water productivity (WP) and water use efficiency (WUE), primarily caused by a lack of sufficient irrigation scheduling technology. To fully address the issue, aiding water scarcity in many regions requires proper technological innovation to manage available resources rather than drawing from new ones.

Thus, my goal became to find solutions that were effective, cost efficient, and had a feasible implementation. I rationalized that if there was a way to know the exact volumetric soil moisture content in the ground at any given time, now or in the future, a decision support system could be utilized to ensure that the proper amount of irrigation was delivered to any crop, precipitating the need for a soil moisture prediction model.

Approaching this problem resulted in a crossroads of several disciplines, including environmental science, mathematics, and computer science. To create this prediction model, I utilized deep learning, namely a recurrent neural network known as an LSTM. I had some primary experience with machine learning and some more basic algorithms, but undertaking this project was my biggest challenge yet. Through online courses, blogs, journals, and discussion forums, I learned the necessary algorithmic components, both theoretical and implemented. At the same time as conducting this project and learning new mathematical, computational, and statistical principles, I was also studying AP Calculus BC and AP Statistics at school.

Conducting research made me realized how I was able to apply knowledge I previously learned, but more importantly, I realized how much I didn't know. It was incredibly intimidating (and sometimes frustrating) to be exposed to so many new ideas and topics at once, but with passion and interest, I was excited at the prospect of applying both new and old interdisciplinary knowledge in the real world.

As for giving some advice, one thing to remember is that there is no easy journey from point A to point B, and research is not a linear path. Each day, there will be something new, and a large aspect of research is embracing both the good and bad parts. Learning from past mistakes, celebrating small victories or steps in the right direction, and most importantly, knowing when to take a step back or try a different approach are some lessons every researcher continuously experiences and learns from for themself. While this non-linearity may fuel ambiguity and hopelessness at times, it is also incredibly rewarding to see all your work come together at the end. I fell in love with research because I viewed it as a way to answer my questions and mitigate uncertainties. I would say I now have many more questions than I began with, but each day, I look forward to exploring new ideas and learning something I didn't know before.

Many students think that every research project needs to be some Nobel Prize level worth endeavor, but everyone must start somewhere. When conducting research, no matter the topic or complexity, every experience is something to learn from. As my school research teacher used to remind us, when starting out, it is important to remember that every master was once a beginner. The best way to start is to explore. My advice to any beginning researcher would be to stay curious, develop your knowledge base, pursue your passions, and don't be afraid about the result. There is no way to guarantee a finding, but by trying, no matter what ends up happening, you end up learning something, and that my friends, is the best feeling in the world.

Research Section

1. Introduction

1.1. Environmental Impacts of Irrigation

Productive irrigation systems are imperative to stimulate economic growth, provide an adequate world food supply by combating food insecurity, and mitigate current problems of water scarcity. However, due to a disparity in technological irrigation advances, many countries and agricultural lands fail to use these practices as adequate means of sustenance. Irrigation is currently the largest water user and waster, utilizing 70-95% of all available water, depending on the region (Mitra et al. 2017; Karmi 2019). Current management and development of irrigation technologies are insufficient, leading to overuse of water for crop production and having a low water use efficiency (WUE), one of the main causes of water scarcity (Mitra et al. 2017). Despite many preconceived notions, fresh water is not an abundant resource on Earth. Currently, 97% of all available water on Earth lies in oceans and seas, 2% lies frozen in glaciers, and only 1% is actually usable for human activity (Stockle, 2001). According to the World Economic Forum, water scarcity is currently projected to be one of the most prominent problems in the future ("Water Scarcity," 2019).

The lack of available freshwater furthers the need for proper irrigation management practices and may potentially reduce the world food supply. Currently, 850 million people suffer from food insecurity, and 1.6 billion tons of food is lost or wasted each year, worth an estimated $1.2 trillion (Karmi, 2019). Problems in harvesting a proper yield are not due to a lack of land available to cultivate, but rather how existing agricultural plots are managed. Thus, most experts believe improving the water use efficiency (WUE) of current agricultural systems is more effective than abandoning existing farmland and beginning new irrigation projects (Stockle, 2001).

Additionally, climate change is projected to increase variability in rainfall, with more dry spells, droughts, and floods, further reducing rainfall and increasing the threat of a lack of fresh water (Rockstrom et al., 2007). However, there is a large disparity in which countries can have easy access to fresh water, and many nations must rely on already dwindling bodies of water such as rivers to harvest freshwater, which can be unreliable and unsustainable (Speck 2020).

1.2. Irrigation Systems Inefficiencies

Traditional methods of irrigation have caused poor management practices, leading to salinization, waterlogging, runoff, chemical water pollution, and public health crisis (Stockle, 2001). Polluted water is a major cause of human disease, and according to the World Health Organization (WHO), more than 4 million children die every year because of water-borne infection. Developments in agricultural technology have shifted from traditional inefficient methods to practices of micro and precision irrigation. Precision irrigation has been shown to improve water use efficiency, reduce energy consumption, and enhance crop productivity by leveraging advances in sensor, control, and modeling technologies (Adeyemi, 2018b). Changing irrigation systems can lead to many effects on an irrigated farm. More efficient irrigation systems make water more productive, yielding more agricultural products than before. However, this could shift agricultural goals to maximizing crop yield, but with the current systems, this will also increase water use. Attempting to weigh technological advancement and economic growth is a delicate balance, as developing efficient irrigation systems can reduce the water required to obtain the same products and maintain the current food supply, but there are also higher water application costs associated with upgrading irrigation technologies (Gomez, 2015). Thus, a costbenefit analysis needs to be undertaken to demonstrate whether upgrading the current irrigation method is necessary.

However, there are other ways of improving WUE and water productivity (WP), that do not involve fully changing a current irrigation system, such as improving irrigation scheduling practices. Irrigation scheduling is a method that determines when and how much water needs to be applied to meet a specific goal (generally to prevent yield-limiting crop water stress) and is essential for sustainable water management and to have profitability optimization in an irrigated farm (Aguilar et al., 2015). Improving irrigation schedules require prediction models that take many factors into consideration, including the type of crop, stage of development, soil properties, soil-water relationships, availability of water supply, and weather conditions (Aguilar et al., 2015). Soil moisture is one of the largest factors when determining how much water an irrigated field needs (Allen, 2020). The most accurate measurement of soil moisture is volumetric water content (volume of liquid water per volume of soil). This metric is then compared with soil water content at field capacity to calculate soil moisture depletion (Sharma and Nelson, 2019).

1.3. Machine Learning Prediction Models and Validation

In creating soil moisture prediction models, various methods have been applied. A Decision Support System (DSS) model uses automated methods to analyze a large amount of unstructured data and accumulate values to aid in decision-making (Corporate Finance Institute, 2020). Existing DSS models can quantify simulated growth values, development, and yield while also providing information about aspects not directly related to crop water needs. However, they usually require a high number of inputs and parameters, and their complexity generally hinders their functionality when applied in the real world, limiting the scope for their implementation (Ramírez-Cuesta et al., 2019). When used for agricultural purposes, DSSs often increase the frequency of irrigation in their schedule but decrease the water used at each interval, lowering overall water use while also relieving crop water stress in a timely manner (Chen et al., 2019).

Prediction models utilize machine learning for creating an automated decision-making system. Machine learning is a data analysis technique that trains computational 'machines' to make predictions on new data. In recent years, there have been several studies exploring the intersections between machine learning and irrigation scheduling, particularly in predictive analysis of soil metrics. Automated machine learning decision support systems that integrate climatic and soil moisture measurements can be used to create irrigation schedules (Adeyemi, 2018b). With this method, different predictor variables have been utilized, along with different model types.

Crane-Droesch used parametric, semi-parametric, and non-parametric models and through a comparison of traditional statistical and new machine learning methods, crop yield values were estimated using weather data. Giusti and Marsili-Libelli utilized a fuzzy decision system in predicting the volumetric soil moisture content based on local weather data. Upadhya and Mathew used fuzzy logic to estimate the crop yield of a given irrigation field, determining the optimized conditions to generate ideal values. Andrade et al. trained Artificial Neural Networks (ANNs) and integrated their function with the Irrigation Scheduling Supervisory Control and Data Acquisition System (ISSCDAS), evaluating the extent to which machine learning predictions can be used to manage an irrigation system autonomously. Adeyemi et al. used Long Short-Term Memory Networks (LSTMs), a class of Recurrent Neural Networks (RNNs) to create a dynamic neural network modeling system that could accurately predict the volumetric soil moisture content of three different sites of potato-growing farms in the United Kingdom with different soil contents based on past soil moisture, precipitation, and climatic measurements. These results were then compared with traditional projections from Forward Feed Neural Networks (FFNNs), a class of ANNs to determine which deep learning technique was superior.

Many studies also validated their predictions either algorithmically using computer-generated simulations to compare automated irrigation schedules with traditional irrigation practices, or experimentally by testing new irrigation schedules and determining if their implementation decreases crop water stress and improves water productivity. Adeyemi et al. used AQUACROP, a software developed by the Food and Agriculture Organization to simulate soil moisture dynamics and crop response to water deficits across the three different locations and soil types. Cao et al. applied short-term weather forecasts to predict precipitation to calculate water deficits that could be used in Alternate wetting and drying (AWD) irrigation, a common practice in rice paddy cultivation. These results were experimentally validated using different methods of irrigation applied across various irrigation fields to determine if irrigation scheduling to reduce water stress would be a viable option for rice paddy farming.

1.4. Purpose

There is an ongoing global problem with a lack of water, due to changes in the climate and poor water management practices. Despite developments and improvements in agriculture technology, WUE, and WP are not improving at similar rates. Ongoing climate change is projected to increase variability in rainfall, with more dry spells, droughts, and floods, further reducing rainfall and increasing the threat of a lack of fresh water, which further limits available resources (Rockstorm et al., 2007). This problem has negatively impacted the world food supply because of mass water wastage which significantly reduces crop yields. This also exacerbates rural poverty as 70% of low-income individuals who have a lack of access to food live in rural areas, meaning these areas require better agricultural practices (Rockstorm et al., 2007). In addition, reducing crop yields will further limit available food sources, increasing food insecurity. A possible cause of this problem is a lack of automated and efficient irrigation scheduling, as most currently agricultural systems use outdated and traditional methods of irrigation.

The purpose of this project is to create an irrigation scheduling model using machine learning methods that can accurately predict soil moisture using various environmental factors and then validate these predictions using algorithmic testing. Some studies have employed similar methods in Europe (Adeyemi, 2018b) and irrigation effects have been vastly studied in Asia as well (Cao et al., 2019), but it is understudied in California, which provides the largest agricultural output in the United States and is an integral part of the world food supply. This study will determine if irrigation scheduling methods are beneficial using predictions and testing in this region. This model should be able to accurately predict soil moisture based on other environmental factors and can create irrigation scheduling systems that improve water use efficiency, fulfilling the project engineering goals.

2. Methods

2.1. Procedures

2.1.1. Data Collection and Study Sites

The data was obtained from a soil moisture probe located at the Desert Chaparral in the University of California, Irvine. This data was obtained through a public access dataset, published in the University of Arizona's Cosmic-Ray Soil Moisture Observing System (COSMOS) monitoring project, and selected because it provided near real-time irrigated field metrics and climatic data for use in a variety of applications including agriculture and water resources management. The data obtained included measurements of rainfall, average daily air temperature, daily maximum air temperature, daily minimum air temperature, and volumetric soil moisture content. The inputs and output values were used along with an optimization process called feature engineering, which involved new variables, “features,” being extrapolated from existing inputs.

2.1.2. Data Cleaning and Preprocessing

To be able to create a machine learning prediction model, the data first had to be transformed into a machine-readable format. This was accomplished by cleaning and standardizing the data values so they could be compared. The hourly data was resampled (recalculated and numerically transformed) to daily intervals. Climatic variables were also resampled, with their respective daily averages calculated, and daily precipitation amounts were recalculated to be the sum of irrigated water depth and average daily rainfall. The volumetric soil moisture content was also resampled to its average daily value, done by combining and averaging its various measurements throughout the day. These values were transformed based on model guidelines to minimize irrigated water use while maximizing water productivity, as proposed to guarantee optimal irrigation operation (Delgoda et al., 2016). The methods proposed by Delgoda et al. were cross validated by Adeyemi et al., ensuring this is a credible process that can be replicated and retested.

The data cleaning process included the inputting of missing (null) values. Additionally, data mined from different sources could include false entries or mistakes, so one way of determining if the values are acceptable is only including values in a certain range, ensuring a false value (or potential outlier) does not cause inaccurate information to be included in the prediction model. This was done to ensure that the data used to algorithmically train the model was accurate and reliable. Additionally, for the prediction model, the data was standardized by calculating the z-scores by firstly converting all values into arrays, making the transformations using a scaler function, and then restructuring the data back into the data frame for analysis. This process was necessary to ensure that differences in ranges of quantities (due to different measurement units) did not affect model predictions.

For the model creation, the dataset was divided into an 80:20 ratio, with 80% of the entire dataset used to train the model, and 20% used in testing. Testing data was used to compare actual values with predicted model values and could not be included in the training set to prevent overfitting.

2.1.3. Recurrent Neural Network Model: Breakdown and Framework

The model was developed using the Keras Deep Learning library in the Python programming language using Jupyter Notebooks. For predictive irrigation scheduling, a one-day-ahead prediction of the soil moisture content was generated. This model was considered a Multiple Input and Single Output (MISO) System because soil moisture depends on the historical and present climate, precipitation, and soil moisture values.

Traditionally, RNNs worked on present input values by considering feedback from previous outputs and storing the memory for a short-term period. However, since this structure fails to store information for a longer period, it is not as useful when long-term information is required to predict current output values. Thus, Long Short Term Memory (LSTM), a class of RNNs was used, which was useful in studying concerning historical data time series analysis (Hochreiter and Schmidhuber, 1997).

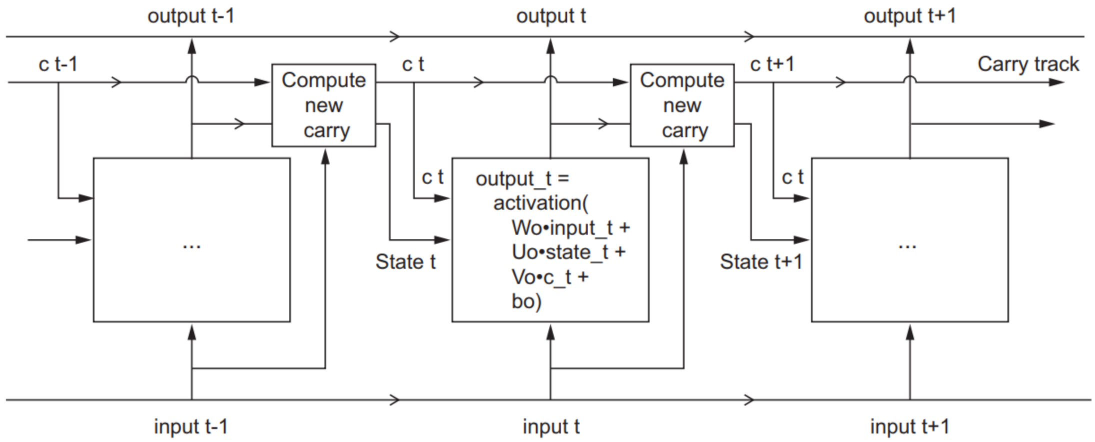

Figure 6.15 Anatomy of an `LSTM`

Figure 1: Diagram showing the basic architecture of an LSTM model (Chollet, 2017)

As indicated by the diagram above, LSTMs predict the next work in each sequence based on previous information (input values). For this project, the final predicted output value (t+1) was the volumetric soil moisture content. LSTMs have 3 different gates and weight vectors; the forget gate disregards information; the input gate handles all current inputs; and the output gate generates predictions at each time stage. Additionally, LSTMs are built with different layers. The "embedding" layer encodes label information into a vector, allowing similar labels with similar vectors to be clustered together. `LSTM` cell layers with `dropout` are where the neural network randomly drops some neurons, meaning one part of the model was sampled and trained in one iteration, and a different part was sampled in another iteration. The network loss was minimized through a statistical technique called Adaptive Moment Estimation (ADAM), which is based on the principles of Gradient Descent. Gradient Descent is an optimization algorithm used to iteratively compute the minimum of a function. With this method, ADAM prevented model overfitting by improving model performance methods to generate more accurate soil moisture predictions.

2.1.4. Model Algorithmic Evaluation

Once the LSTM model was created, it was used to make t+1 soil moisture predictions. To validate the accuracy of these predictions, residual plots of predicted and actual soil moisture values (from the testing data) were created to determine the error. Then, the mean squared error (MSE) of all the predicted values was calculated, with potential values from [0,), and a value closer to 0 is indicative of a stronger model. The coefficient of determination (R^2 value) was also calculated, with potential values from 0,1, and a value closer to 1 is indicative of a stronger model. The predicted and actual values were also compared using a t-test to determine if the results were statistically not significant because the expected outcome searched for similarity between real and predicted values.

3. Results and Analysis

3.1 Test Site Evaluations

Desert Chaparral UCI was the evaluation site selected for this project because Southern California remains a prevalent region in terms of the world agricultural supply. Initially, the process began with the notion that the designated area of interest would be in the American Midwest, with analysis across several states in this region, modeling the ideas after the study conducted by Adeyemi et al. However, certain limitations within the procedures existed, including a sheer lack of data and disorganized records, which hindered the ability to data-mine and select those locations. As indicated by the sample table above, the data coverage available for midwestern states was suboptimal, and since California had the largest agricultural yield, the experiment was modified to concentrate on several test sites in this state, as opposed to testing sites in several different states.

3.2 Aggregated Dataset- Training Data Optimization

Upon selection of the test site, multiple data sets were downloaded as Comma Separated Values (CSV), and the cleaning process involved resampling the data to have the date set as an index value. Additionally, there were discrepancies between the data for COSMOS and NOAA. Besides the different scales of measure (which did not affect the final model as the data was standardized), there were discrepancies between the frequencies in which the data was recorded. Soil moisture readings were recorded several times throughout the day, whereas the weather data records available only included daily readings. Thus, the data was analyzed by computing the mean daily soil moisture by grouping data values that had the same index value (date). Due to a lack of proper data and the subsequent cleaning process that ensued, parts of the initial data had to be completely disregarded, with other categories being heavily cut down; while 65,535 data points worth of readings were initially collected, this number was cut down to 2,902 values of clean, organized data with no null or incorrect numbers. This was another potential limitation because while this data was adequate for creating and testing a machine learning model, larger datasets are always desired to improve model performance, which may have hindered model prediction capability to a certain extent.

Table 2: Sample (pre-standardized) aggregated dataset with soil moisture, maximum, minimum, average temperature, and precipitation using date as index value

	SOILM	TMAX	TMIN	PRCP	TAVG
DATE					
2011-08-31	3.060000	93.0	59.0	0.0	76.00
2011-09-01	3.275000	94.0	59.0	0.0	76.50
2011-09-02	3.787500	94.0	61.0	0.0	77.50
2011-09-03	3.916667	95.0	61.0	0.0	78.00
2011-09-04	3.895833	94.5	61.0	0.0	77.75

Table 2 depicts the various input and output values. The final aggregated dataset was standardized into z-scores by coding a data transformer using a RobustScaler function (from the ScikitLearn Python Library) and rearranging the data into arrays for analysis.

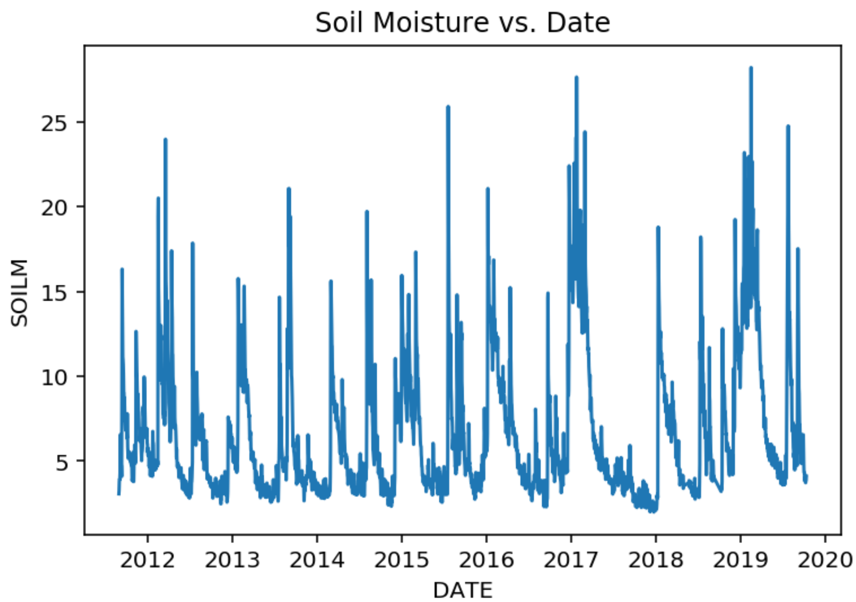

Figure 2: Line plot of soil moisture readings vs. date (created using Seaborn Python library)

Figure 2 highlights a distinct cyclic pattern with the soil moisture based throughout a 9-year time period (2011-2019). While the values have slight fluctuations, distinct trends emerge through the line plot, depicting how soil moisture readings predictably change throughout a year, depending on the temperature, precipitation, growing season timings, etc. This figure emphasized the importance of selecting the proposed machine learning model: a time series analysis. For prediction, the time of year is vastly important, and thus the data cannot be scrambled or randomized, because the neural network needs to take timing into consideration.

3.3 LSTM Model Creation and Algorithmic Evaluation

For the model training process, hyperparameters were tested and optimized. The number of neurons, which affects the learning capacity of the network, was selected as 1. Typically, more neurons would be able to learn more structure from the problem but increasing this number will severely increase training time. This increased learning capacity may also potentially overfit the training data. The batch size, which controls the number of data subsamples used, was set at 32, tailored to fit the relative size of the dataset. The time step value was set at 30, meaning that 30 previous data values (about a month) would be considered in the prediction of the *t+1* desired output. The data subsequences were not shuffled to retain the time series analysis as *time-of-year* was an important aspect of creating the model.

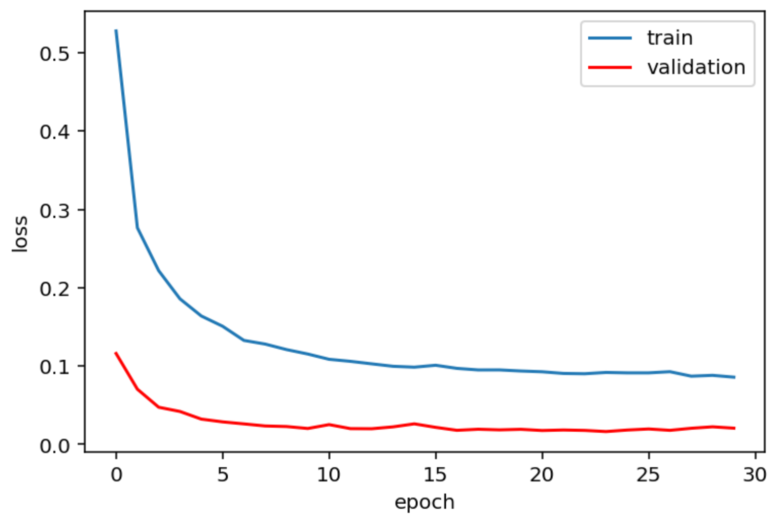

Figure 3: Plot of the training loss and validation loss as the epoch number approached maximum value (30)

The validation losses were computed for the model to diagnose whether that fit was adequate for prediction, or whether the models were overfitted/under-fitted. Generally, if a model has a low train accuracy and a high train loss, then the model is under-fitted and if the model has a high train accuracy but a low validation accuracy then the model is overfitted. Figure 3 shows a plot of the training and validation losses, which are at relatively low values, and approach each other as the epoch value approaches 30. Thus, the model was well fit for the dataset.

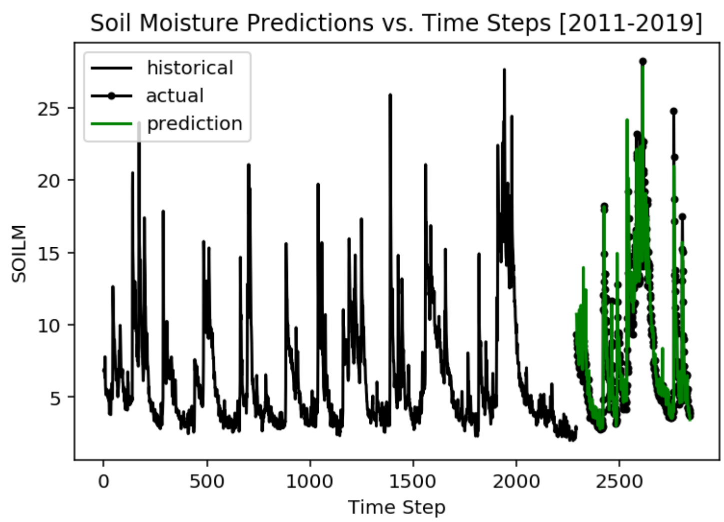

Figure 4: Historical data of Soil Moisture readings vs. Time Steps plotted for time period, with predicted values

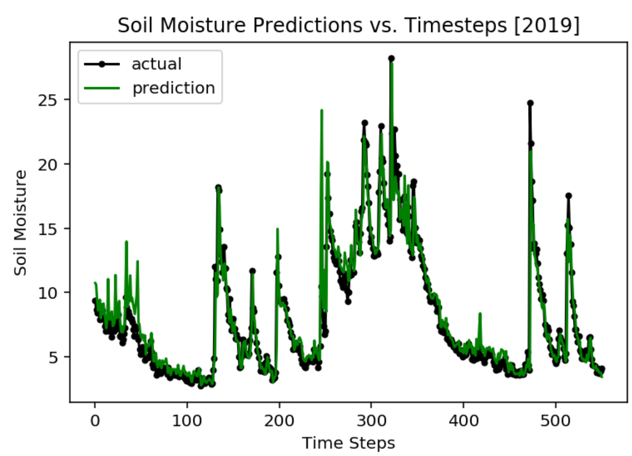

Figure 5: Historical data of predicted and actual Soil Moisture readings vs. Time Steps plotted for testing data period

Figures 4 and 5 depict the predicted (green) and actual (black) soil moisture models. These plots indicate the model was able to provide a good fit for the data and adhere to match its unique cyclic patterns. While this model was a good fit, as indicated by this plot, the computer used did not have enough RAM for heavy machine learning processing, so while the hyperparameters were optimized to the largest scope possible, there could be potential for an even more accurate model in the future.

3.4 Model Statistical Evaluation

Certain statistical tests were performed to further validate these results. Firstly, the mean squared error (MSE) was computed to be 0.213; this low value is indicative of an accurate model. The r^2 value was also computed to be 0.852, which indicates that about 85% of the variation in the actual soil moisture values can be explained by the predicted y-values. A t-test was also run between the actual and predicted soil moisture values to determine if the results could be deemed significant. The resulting p-value was 0.47, greater than an alpha value of 0.05, meaning there is no statistically significant difference between the actual and predicted results, demonstrating model accuracy.

When compared to the study by Adeyemi et al. in England, despite vast differences in geographic location and soil profiles, the study's findings matched their results as the LSTM algorithmic structure was successful. The model choice was validated by Ayedemi et al. through their demonstration that between Forward Feed Neural Networks, LSTMS, and traditional statistical models, significant evidence remained in favor of the applications of LSTM models in soil moisture predictions and irrigation scheduling (Adeyemi et al., 2018b). Additionally, these predictions were successfully algorithmically and statistically verified, with a large r^2 value and small MSE, like Adeyemi et al and Crane-Droesch.

4. Conclusion

4.1 Summary of Findings

In conclusion, the results of this project indicate the completion of the project's engineering goals and fulfillment of the gap. The results of this study validate the completion of the project's engineering goals and address a gap in the scholarly conversation. Upon conducting a t-test, a p-value of 0.47 was obtained, greater than the alpha of 0.05, showing that there was no statistically significant difference between the actual soil moisture values and the model's predicted values. This was an algorithmic and statistical verification of the efficacy of the model, demonstrating the use of LSTMs to create a prediction model. The results also demonstrate filling in the gap since California, which has the highest agricultural production in the US and is a major source of the world food supply, was heavily understudied. With this new information, the principles of optimizing certain hyperparameters, and the knowledge gained in terms of fine-tuning a machine learning model were greatly increased, which may create better soil moisture models and irrigation scheduling systems in general.

4.2 Limitations

One of the limitations of the data was the small region that was primarily focused on. Since only Southern California between Los Angeles and San Diego was considered, the observations generated findings for a small region, so while this model and its complexities influenced the understanding of soil moisture prediction capabilities in this region, this information is not necessarily applicable to the rest of the world. Additionally, there are many factors involved in creating a model, firstly in the crop yield parameters such as crop and soil type. Weather conditions, the accuracy of soil moisture probes, and access to clean and accurate data, all can change depending on the region of interest, so if implemented, soil moisture prediction models will have to be fine-tuned by territory or region of interest. Additionally, due to the limited RAM available on the computer used, the ability to test certain hyperparameters and optimize the model was hindered, so an LSTM model can be fine-tuned to generate even more accurate predictions. LSTM was also the only algorithm that was used in data modeling because it was heavily researched that this would be optimal for data type and type of predictions, but since other deep learning models or algorithms were not tested, there is always a possibility that more accurate soil moisture prediction models can be created using another decision support system or machine learning method.

4.3 Implications

This project utilizes machine learning algorithms to create soil moisture, prediction models. Since a comprehensive soil moisture model takes precipitation, groundwater, temperature, etc. into account and can predict the soil moisture for any given day in a year, this prediction can be used in irrigation scheduling. By knowing how much water is currently in the soil, we can use optimization equations based on the type of crop and soil to determine the maximum volumetric soil moisture content, and then determine if irrigation needs to be applied in a given day. The soil moisture model is the first step to creating a highly successful irrigation decision support system and enables an agriculturalist to conserve water by applying the exact, optimized amount of water. Predicting soil moisture also improves the water productivity in a system, since a crop is less likely to be overwatered or underwatered. These findings also demonstrate the importance of predictions in irrigation scheduling. While weather stations and soil moisture probes are currently used by agriculturalists to manually schedule irrigation for their crops, predictive models can create a more automated system, and as the accessibility of such technology expands, producing crops will become more efficient.

4.4 Future Directions

These findings precipitate several future projects to gain a better understanding of this field. Primarily, one of the largest shortcomings within the project was the lack of varied geographic locations for which data was available. Creating soil moisture prediction models is essential for irrigation scheduling, but there are many factors within a given region that can affect soil moisture, and while some climatic conditions were studied, such as temperature, precipitation, etc., ultimately a lack of data parameters contributed to fewer factors being observed. Future studies could aim to access more types of data with more climatic variables—such as snow cover, sunlight, humidity, and evapotranspiration—as well as other irrigation field observations, such as soil type, crop type, and irrigation method. Ultimately, adding more input variables may heighten the accuracy of the model, provided measures are put into place to optimize hyperparameters without overfitting. Future studies could also investigate different regions, instead of one concentrated area, or investigate a specific irrigated field type. Overall, the success of any such projects ends up depending on the type, quality, and quantity of data available, so in the future, while public access data sources are sufficient, to gain better and more comprehensive access, work can be conducted in an irrigated field or local academic institutions. Additionally, future studies can look into creating extremely accurate models and then creating an automated decision support system built into a usable application for agriculturalists and farmers. This will enable farmers to tailor predictions to their own needs, and determine how much water must be applied, aiding in water conservation. Finally, another future study can attempt to use the predictions created here and test them in real-time, growing different crops using different irrigation scheduling methods (based on different soil moisture prediction values), in a way to test and confirm not just algorithmically or through simulations, but in a real irrigated field.

5. References

Adeyemi, O., Grove, I., Peets, S., Domun, Y., & Norton, T. (2018b). Dynamic Neural Network Modelling of Soil Moisture Content for Predictive Irrigation Scheduling. *Sensors*, *18*(10), 3408. https://doi.org/10.3390/s18103408

Aguilar, J., Rogers, D., & Kisekka, I. (2015). Irrigation Scheduling Based on Soil Moisture Sensors and Evapotranspiration. *Kansas Agricultural Experiment Station Research Reports*, *1*(5), 1–6. https://doi.org/10.4148/2378-5977.1087

Al-Kaisi, M. M., & Brone, I. (2016, March 7). *Crop Water Use and Growth Stages - 4.715*. Colorado State University Extension. https://extension.colostate.edu/topic-areas

/agriculture/crop-water-use-and-growth-stages-4-715/

Allen, L. (2020, September 1). *Drip – cotton's 'Cadillac of irrigation systems.'* Farm Progress. https://www.farmprogress.com/water/drip-cottons-cadillac-irrigation-systems

Andales, A. A., Chavez, J. L., & Bauder, T. A. (2016, March 8). *Irrigation Scheduling: The Water Balance Approach - 4.707*. Colorado State University Extensions. https://extension.colostate.edu/topic-areas/agriculture/irrigation-scheduling-the-water-balance-approach-4-707/

Cao, J., Tan, J., Cui, Y., & Luo, Y. (2019). Irrigation scheduling of paddy rice using short-term weather forecast data. *Agricultural Water Management*, *213*, 714–723. https://doi.org/10.1016/j.agwat.2018.10.06

Chollet, F. (2017, November). Anatomy of a LSTM [Diagram]. In *Deep Learning with Python*.

Corporate Finance Institute. (2020, July 31). *Decision Support System (DSS)*. https://corporatefinanceinstitute.com/resources/knowledge/other/decision-support-system-dss/

Delgoda, D., Malano, H., Saleem, S. K., & Halgamuge, M. N. (2016). Irrigation control based on model predictive control (MPC): Formulation of theory and validation using weather forecast data and AQUACROP model. *Environmental Modelling & Software*, *78*, 40–53. https://doi.org/10.1016/j.envsoft.2015.12.012

Giusti, E., & Marsili-Libelli, S. (2015). A Fuzzy Decision Support System for irrigation and water conservation in agriculture. *Environmental Modelling & Software*, *63*, 73–86. https://doi.org/10.1016/j.envsoft.2014.09.020

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Karmi, A. (2019, March 26). *How Smarter Technology Will Feed the Planet*. Scientific American Blog Network. https://blogs.scientificamerican.com/observations/how-smarter-technology-will-feed-the-planet/.

Mitra, B. K., Sahin, S., Markandya, A., & Bao, P. N. (2017). *Improving Irrigation Water Use Efficiency Holds the Key to Tackling Water Scarcity in South Asia: Technical Potential and Financing Options*. Institute for Global Environmental Strategies

Sharma, V., & Nelson, A. (2019, April 29). *Soil moisture sensors for irrigation scheduling*. Minnesota Crop News. https://blog-crop-news.extension.umn.edu/2019/04/soil-moisture-sensors-for-irrigation.html

Stockle, C. O. (2001, August). *Environmental Impact of Irrigation: A Review*. State of Washington Water Research Center, Washington State University. https://www.swwrc.wsu.edu/newsletter/fall2001/IrrImpact2.pdf

Rockstrom, J., Lannerstad, M., & Falkenmark, M. (2007). Assessing the water challenge of a new green revolution in developing countries. *Proceedings of the National Academy of Sciences*, *104*(15), 6253–6260. https://doi.org/10.1073/pnas.0605739104

Upadhya, S. M., & Mathew, S. (2020). Implementation of Fuzzy Logic in Estimating Yield of a Vegetable Crop. *Journal of Physics: Conference Series*, *1427*, 012013. https://doi.org/10.1088/1742-6596/1427/1/012013

Water scarcity is one of the greatest challenges of our time. (2019, March 22). World Economic Forum. https://www.weforum.org/agenda/2019/03/water-scarcity-one-of-the-greatest-challenges-of-our-time