Prediction and Mapping of Stem Water Potential in Almond Orchards Using Remote Sensing and Machine Learning

By

PETER SAVCHIK

THESIS

Submitted in partial satisfaction of the requirements for the degree of

Master of Science

in

Biological Systems Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Isaya Kisekka, Chair

Andre Daccache

Mallika Nocco

Committee in Charge

2023

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS……..iv

LIST OF TABLES……v

LIST OF FIGURES……vi

LIST OF ABBREVIATIONS……...vii

ABSTRACT……viii

CHAPTER

1 REVIEW OF LITERATURE AND RATIONALE……..…1

Precision and Regulated Deficit Irrigation……..1

Stem Water Potential……...2

Crop Water Stress Indices……2

Machine Learning for Water Status Prediction……...2

Vegetation Indices and Hyper/Multispectral Indices……...5

Soil Water Potential……8

Hypothesis……9

Objectives……9

2 PREDICTION OF STEM WATER POTENTIAL IN ALMONDS USING REMOTE SENSING AND MACHINE LEARNING BASED MODELS……10

Introduction……10

Materials and Methods……...13

Study Site and Experimental Layout……13

Soil Moisture Measurements……14

Midday Stem Water Potential Measurements……14

Multispectral Imagery Collection……..15

Supplemental Environmental Data……16

Processing of Multispectral Imagery……16

Artificial Neural Network Training and Sensitivity Analysis……...16

Stem Water Potential Mapping Using Artificial Neural Network Model……17
Random Forest Forward Feature Selection……18
Model Training and Target Oriented Validation……...20
Stem Water Potential Mapping Using Random Forest Model……..21
Results..……..21
Single Season Artificial Neural Network Modeling……..21
Final Artificial Neural Network Modeling……21
Artificial Neural Network Stem Water Potential Map Performance……23
Random Forest Variable Importance Analysis……..23
Random Forest Feed Forward Selection……24
Random Forest Modeling and Map Performance……..24
Comparison of Random Forest and Artificial Neural Network Models……25
Comparison of Model and Map Performance……25
Discussion……..33
3 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH……..35
Conclusions……35
Recommendations for Future Research……36
References……..37

ACKNOWLEDGEMENTS

I have many to thank for my success in completing this thesis. First, I would like to thank my committee members Dr. Isaya Kisekka, Dr. Mallika Nocco, and Dr. Andre Daccache for their continued feedback and guidance throughout my master's program. I extend my sincere gratitude to all of my undergraduate assistants who helped with all data collection and many difficult tasks, without whom this project would not have been possible: Jazmin Melendez, Krista Blide, Evan McKenzie, Jacob Mueller, and Aaron Guerra. I would also like to thank the NIFA Artificial Intelligence in Food Systems Institute for their continued funding and support throughout this project.

I would like to thank my parents who have supported and inspired me to pursue science since day one. Most of all, thank you to my partner Samantha, whose love and support has inspired me to push through the most challenging of times in my pursuit of higher education.

LIST OF TABLES

Table page

1 Descriptions of Abbreviations Used……...iv

2 Performative Statistics for all Models……22

3 List of Names and Input Parameters for all Models……..22

4 Summary of Performative Statistics for all Maps……..25

LIST OF FIGURES

Figure page

1 Basic Structure of an Artificial Neural Network…….3

2 Basic Structure of a Random Forest Model…….4

3 Locations of Focus and Validation Measurement Trees Within Project Site……13

4 Images of PMS Model 615 Pressure Bomb and Instrotek 503 Hydroprobe……..15

5 Images of Equipment Used for Multispectral Data Collection……..15

6 Examples of Input Layers Used in Spatial Modeling……18

7 Results of Feed Forward Analysis Applied to Random Forest Model……..19

8 Results of Variable Importance Analysis on all Potential Predictors……19

9 Results of Variable Importance Analysis on Target Oriented Validation…….20

10 Artificial Neural Network Baseline-Adjusted Stem Water Potential Maps……..27

11 Random Forest Baseline-Adjusted Stem Water Potential Maps……...28

12 Artificial Neural Network Raw Stem Water Potential Maps…….29

13 Random Forest Raw Stem Water Potential Maps……..30

14 Map Performance at Data Sampling Locations…….31

15 Map Performance at Validation Locations……32

Table 1: Description of abbreviations used in "Prediction and Mapping of SWP in Almonds Using Remote Sensing and Machine Learning

Abbreviation	**Meaning**
SWP	Stem Water Potential
NP	Neutron Probe
ET	Evapotranspiration
ANN	Artificial Neural Network
RF	Random Forest
NDRE	Normalized Difference Red Edge Index
RMSE	Root Mean Squared Error
nRMSE	Normalized Root Mean Square Error
IOA	Index of Agreement
RDI	Regulated Deficit Irrigation
CWSI	Crop Water Stress Index
FRF	Transpiration Reduction Function

ABSTRACT

Almonds are a major crop in the state of California, in which 90% of all the world's almonds are produced. Widespread drought and strict groundwater regulations pose significant challenges to growers throughout the state. Irrigation regimes based on observed crop water status can help to optimize water use efficiency, but consistent and accurate measurement of water status can prove challenging. In almonds, crop water status is best represented by midday stem water potential, which despite its accuracy is impractical for growers to measure on a regular basis. This study aimed to use machine learning models to predict stem water potential in an almond orchard based on canopy spectral reflectance values, soil moisture, and daily evapotranspiration. Both artificial neural network and random forest models were trained and were used to produce high resolution spatial maps of stem water potential covering the entire orchard area. Additionally, for each model type one model was trained to predict raw stem water potential values, while another was trained to predict baseline-adjusted values. Together, all models resulted in an average coefficient of correlation of R^2=0.73 and an average root mean squared error (RMSE) of 2.5 bar. Prediction accuracy decreased significantly when models were expanded to spatial maps (R^2=0.33, RMSE=3.31 [avg]). These results indicate that both artificial neural network and random forest frameworks can be used effectively to predict and map stem water potential, but that both approaches are unable to fully account for the spatial variability observed throughout the orchard. Random forest models predicting raw stem water potential produced the most accurate maps. Overall, the most accurate maps were those produced by the random forest modeling predicting raw stem water potential values (R^2=0.47, RMSE=2.71 bar).

CHAPTER 1: REVIEW OF LITERATURE AND RATIONALE

Precision and Regulated Deficit Irrigation

Due to the high demand for water use efficiency in the almond industry, precision irrigation techniques are widely used by growers in one form or another. Broadly, precision irrigation refers to an irrigation regime in which water is only applied "in the right place at the right time" to achieve a balance of optimal production and efficiency. Effective precision irrigation requires some knowledge of the water status of plants, preferably with a high spatial and temporal resolution. Across a field, crop water requirements may vary greatly due to differences in canopy size, soil variability, and field topography (Starr, 2005). Requirements may also change throughout the growing season due to changing weather patterns (Goldhamer, 2005). In a 2018 study, application of a precision irrigation regime based on estimated crop water status in almonds resulted in a 39% reduction in water usage compared with a control treatment of 100% evapotranspiration (ET) replacement (Dhillon et al, 2018).

Regulated Deficit Irrigation (RDI) is a type of precision irrigation in which the crop is intentionally stressed at strategic growth stages, and water is only applied once the estimated crop water stress reaches a certain level. Proper RDI application in almonds allows growers to stretch irrigation supplies and reduce crop evapotranspiration, with minimal losses in yield (DeJonge et al, 2015). In 2004, Goldhamer et al found an RDI regime applied to pistachios saved 23.2% of total water relative to the control, and significantly increased the water efficiency of the crop from 3.61 to 4.69 (kg marketable fruit/mm water). However, if conditions are not properly assessed, RDI may risk significant reductions in crop yield and quality. Thus, careful monitoring of water stress is key to successful implementation of RDI (DeJonge et al, 2015).

Stem Water Potential

The most widely accepted method of determining the water status of a plant is to measure its midday stem water potential (SWP) using a pressure chamber (Drechsler, 2019). To do so, a leaf is cut from the crop and placed inside a sealed chamber. Using compressed nitrogen, the pressure in the chamber is then slowly increased until sap comes out of the leaf petiole (Dhillon et al, 2018). Obtaining SWP measurements is tedious and time consuming and is thus not used as a water status prediction tool in commercial settings.

Crop Water Stress Indices

To estimate crop water status without the need for laborious measurements, a number of stress indices have been formulated over the last few decades, which may be used to estimate crop water stress. The most notable of these is the Crop Water Stress Index (CWSI) derived by Jackson et al in 1981. The CWSI is the ratio of a crop's canopy temperature minus the air temperature, relative to the temperature differential of a well-watered plant and a non-transpiring plant (DeJonge et al, 2015). CWSI has been demonstrated on numerous occasions to correlate well with SWP, making it a potentially viable tool for water stress prediction and irrigation decision making. However, it still requires calculation of a well-watered (saturated) baseline (which is tedious in a large commercial setting), and its correlation with SWP often varies throughout the growing season (Möller et al, 2007). While the CWSI has its drawbacks, its consistent correlation with SWP indicate that its key parameters, namely air temperature, leaf/canopy temperature, relative humidity, and vapor pressure deficit may be useful inputs in a predictive model of SWP.

Machine Learning for Water Status Prediction

In recent years, machine learning algorithms such as Artificial Neural Networks (ANN) and Random Forests (RF) have risen in popularity as tools for modeling the complex variables and relationships behind plant water stress.

Artificial neural networks are a type of computational network simulating the learning structure of the brain and nervous system. They can be applied to perform regression analysis of highly non-linear problems and relationships (Hsu et al, 1995). They are comprised of input, hidden, and output layers, with each layer consisting of a number of neurons representing predictive variables. Figure 2 illustrates the structure of an ANN model. The number of hidden neurons may vary and is defined by the user. The output of each neuron is a function of the weighted sum of its inputs. Once input and target values (outputs) have been defined, the network is able to learn the proper weights and biases that will result in accurate prediction of output values (Meyers et al, 2015).

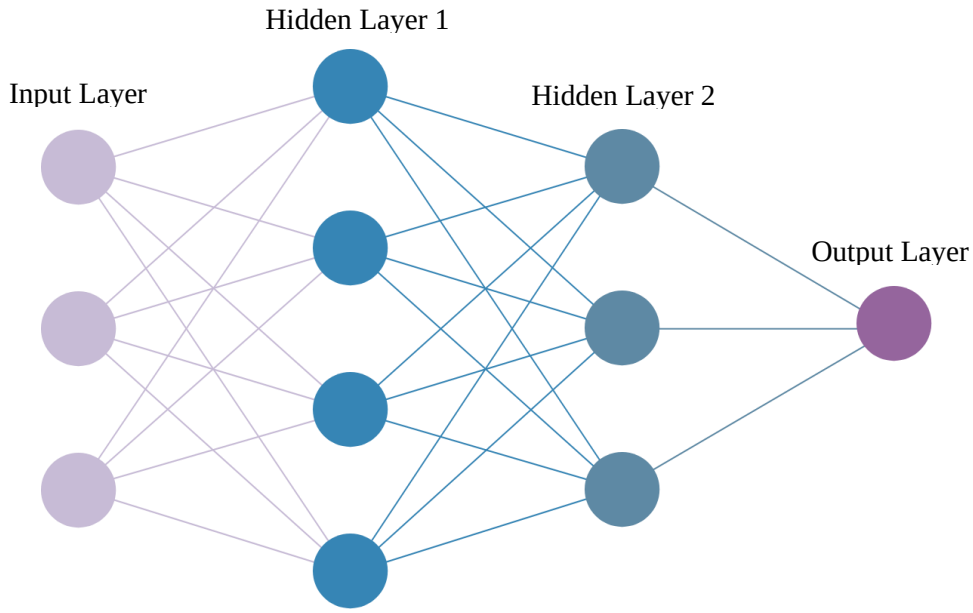


Figure 1) Basic Structure of an Artificial Neural Network

In a random forest model, samples are selected randomly with replacement from a training dataset to form a network of decision "trees", which together form a "forest". The general structure of a random forest can be viewed in figure 2. A key advantage of random forests is that they are highly resistant to overfitting as a result of the high number of decision trees they typically contain (Virdnokar et al, 2020). The model is defined by two main hyper parameters; number of trees (n) and number of features used in decision making at each tree (mtry). Based on resulting error stabilization, Belgiu et al

(2016) determined n=500 as the optimal number of trees for most modeling scenarios. Mtry is typically set to a value of p/3, where p represents the number of predictive variables (Gislason et al, 2006). Alternatively, p/3 may also be used if an exact integer value is preferred.

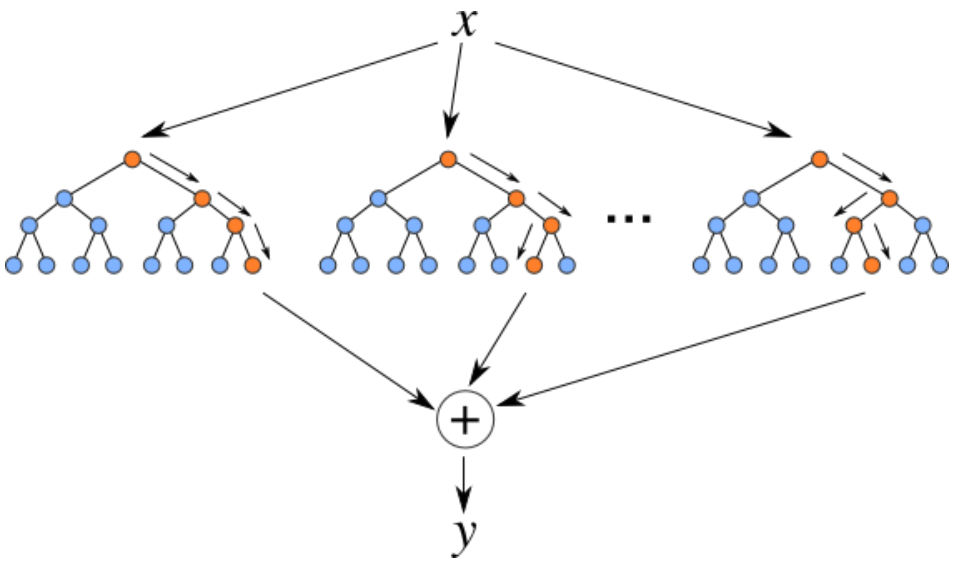


Figure 2) Basic Structure of a Random Forest Model

ANNs have already been utilized in many cases for water stress prediction, most commonly in grapevines (Virdonkar et al, 2020). Utilizing thermal indices, canopy temperature data, and air temperature, Gutierrez et al (2018) programmed an ANN to predict SWP in grapevines (R^2=0.61). It was found that whether baseline temperatures for the indices were included as inputs or not, the ANN predicted SWP equally well. In another vineyard study, Poblete et al (2017) programmed an ANN to predict SWP using different combinations of multispectral imagery, and the resultant models predicted SWP with moderate to high accuracy, producing coefficients of correlation from R^2=0.58-0.87 (Poblete et al, 2017). Similarly, Romero et al predicted SWP across a vineyard with good accuracy, utilizing ten vegetation indices derived from multispectral data as ANN inputs (Romero et al, 2018). In another case, Marti et al used both an optimum regression equation and an ANN model to predict SWP in fruit trees, with temperature data, relative humidity, solar radiation, and soil moisture as inputs. The ANN model

resulted in an R^2 value of 0.93 between predicted and observed values, while the regression equation resulted in an R^2 value of 0.85, indicating the superiority of an ANN model for water status prediction. ANN applications to water status prediction in almonds thus far are somewhat limited and have mostly been carried out in a "proof of concept" fashion. Meyers et al used an ANN to predict an alternative plant water status indicator in almonds, $T_{diffdry}$, interpreted as the difference between the temperature of a dry leaf and a live leaf. An ANN was trained several different times, using varying combinations of input parameters. The most successful model utilized SWP, soil water potential, vapor pressure deficit, and photosynthetically active radiation as inputs, and resulted in an R^2 value of 0.78 between predicted and calculated values of $T_{diffdry}$.

Compared to ANNs, applications of RF models to water stress prediction are notably fewer in number (Virdnokar et al, 2020). Using various meteorological data, Yang et al (2021) used a random forest model to predict canopy temperatures of Chinese Brassica (R^2=0.77-0.9), the values of which were subsequently used for CWSI calculation. In a 2018 study based in South Africa, a random forest classification model was used to determine the level of water stress in grape vines, with a test accuracy of 83.3% (Loggenberg et al, 2018). In another case, Pocas et al (2017) investigated the ability of nine different types of regression models, including random forest, to predict vineyard pre-dawn SWP in the Douro wine region of Portugal. Each model utilized various hyperspectral vegetation indices as predictive variables. The random forest model predicted vineyard SWP with an average R^2 value of 0.77 and an average RMSE of 0.11 MPa, indicating a promising potential for the application of random forest models to water status prediction.

Vegetation Indices and Hyper/Multispectral Imagery

Reflectance values of specific spectral bands throughout the electromagnetic spectrum data has been known for some time to relate to plant water content, though the specific nature of that relationship has remained rather ambiguous. What is known is that water absorbs radiant energy from many wavelengths throughout the electromagnetic spectrum, particularly throughout the mid-infrared region

(1300-2500 nm), with significant absorption bands centered on 1450, 1940, and 2500 nm wavelengths (Oumar et al, 2013). By extension, absorption of 400-2500 nm radiation by water causes the reflectance of a plant's leaves to decrease. When water is lost from leaves, there is an increase in intracellular air space, consequently increasing the intensity of reflections within the leaves. This phenomenon can partially explain how water loss may increase reflectance throughout certain regions of the electromagnetic spectrum (Carter, 1991).

Similar to crop water stress indices, many vegetation indices (VI) exist which can be used to estimate crop water status (Berni et al, 2017), among other applications such as yield prediction, disease detection, and phenotyping (Parker et al, 2020). Vegetation indices are calculated from remotely sensed multispectral imagery and are based on ratios of the reflectance of specific spectral bands. The most widely used vegetation index is the Normalized Difference Vegetation Index (NDVI), calculated using the following equation (Berni et al, 2017):

$$NDVI = (R_{800} - R_{670}) / (R_{800} + R_{670}) \quad (1)$$

Where R_{800} is the reflectance of the 800nm near-infrared (NIR) spectral band, and R_{670} is the reflectance of the 670nm green spectral band. NDVI has been proven to correlate to some degree with water stress, however it is generally considered to reflect more of a cumulative long-term response to water stress in crops rather than immediate effects of stress (Baluja et al, 2012). Another VI is the Photochemical Reflectance Index (PRI) calculated with the following equation (Berni et al, 2017):

$$PRI = (R_{570} - R_{531}) / (R_{570} - R_{531}) \quad (2)$$

Where R_{570} is the reflectance of the 570nm spectral band, and R_{531} is the reflectance of the 531nm spectral band. In a 2017 study, PRI was found by Berni et al to correlate with canopy temperature in maize (R^2=0.69), indicating potential viability as a predictor of water stress. A separate study by Berni et al (2009) the visible, red edge and near-infrared regions to correlate best with water stress in canopy crops. As PRI is comprised of only visible spectral bands, a VI that encompasses both the red edge and infrared

regions may prove more effective for water stress estimation in crops such as almonds. One such VI is the Normalized Difference Red Edge Index (NDRE), obtained using the following equation:

$$NDRE = (R_{800} - R_{720}) / (R_{800} + R_{720}) \quad (3)$$

Where R_{800} is the reflectance of the 800nm NIR spectral band, and R_{720} is the reflectance of the 720nm red edge spectral band. Zhang et al (2019) found NDRE to vary significantly in response to changes in canopy water content of summer maize. The study also found that canopy reflectance of the red edge spectral band itself correlated well with canopy water content (R^2=0.77-0.79).

In recent years, VIs and multispectral data have shown promise for predicting water stress in different crops. Romero et al (2018) used ten different VIs as inputs to an ANN to predict SWP across a vineyard. While all the indices individually had very low correlations with SWP, the ANN using them all as inputs produced a high coefficient of correlation between predicted and observed SWP (R^2=0.8 with training data, R^2=0.72 with validation data). A new dataset was obtained from a different vineyard site and input into the model, again producing a high correlation (R^2=0.83). All the indices used in this study were calculated using some numerical combination of the near-infrared, red, and green multispectral bands' reflectance. While none of the indices on their own demonstrated strong prediction of water status, the results of this study seem to indicate that the indices (or the multispectral bands used to calculate them) may be good predictors of water status in combination with other parameters in regression model.

In addition to visible and near-infrared imagery, other regions of the electromagnetic spectrum have shown promise as well for water status prediction. In a 2013 study, Oumar et al programed a neural network to predict plant water content (PWC) in trees throughout a eucalyptus plantation, utilizing a number of bands and spectral indices from the 800-2000 nm range. The model predicted plant water content with a correlation of R^2=0.88. A sensitivity analysis of each input variable was conducted, and revealed that 1450 nm reflectance, Normalized Difference Water Index (NDWI), Normalized Difference

Infrared Index (NDII), and 1200 nm reflectance (in that order) were the most important variables to the model's prediction. Regardless of order, all four parameters were found to be significantly more important to the model's predictions than any other input variables. NDWI is the normalized difference of 860 and 1240 nm band reflectance, while NDII is the normalized difference of 819 and 1649 nm band reflectance. Results from this study indicate that 1450 and 1200 nm band reflectance, as well as the reflectance of the bands comprising the NDWI and NDII may be good predictors of SWP in an model alongside other input data. However, a major drawback of measuring reflectance for wavelengths larger than 800nm is the cost of equipment. Cameras with the ability to measure higher spectral ranges are far less common and significantly more expensive than those which measure the visible and NIR regions. Thus, reflectance of bands beyond 800nm were not included in this study.

Soil Water Potential

Soil water status has long been considered an important factor in irrigation planning (Campbell et al, 1982). While soil water status is not always equivalent to a plant's water status, especially in orchard crops with deep root systems (Naor, 2008), it can serve as a guiding factor for water status indication. A transpiration reduction function (FRF) can be used to correlate soil water potential to crop water status. Similar in concept to CWSI, FRF uses a ratio of soil water potential differentials to estimate relative transpiration. Durigon et al found good correlations in common bean when comparing FRF derived relative transpiration and CWSI (R^2=0.64-0.67), implying soil water status can be a useful predictive variable in modeling of plant water status.

Hypothesis

The literature has shown clearly that RDI implementation may help almond growers produce steady yields with less water, provided they can accurately assess crop water status. Machine learning models have been demonstrated to have the potential to produce accurate and timely water stress approximations. Remotely sensed multispectral data may serve as useful predictive variables in such a model, along with other soil, plant, and climatic variables. This study puts forward two hypotheses:

1) With the right input variables, a machine learning model can accurately predict SWP in almond trees.
2) Such a model can accurately map spatial SWP throughout an almond orchard.

Objectives

The overall objective of this study is to map water stress in almond orchards using machine learning based models. The specific objectives are as follows:

1) Develop machine learning based models for SWP prediction in almond orchards.
2) Use selected models to produce high resolution spatial maps of predicted SWP across the orchard.
3) Compare the performance of ANN and RF based maps.

CHAPTER 2: PREDICTION OF SWP IN ALMONDS USING REMOTE SENSING AND MACHINE LEARNING

Introduction

Almonds are a major crop in the state of California, contributing an estimated $11 billion annually to the state's GDP (Sumner, 2014). Along with their profitability, almonds are a highly water-intensive crop, requiring year-round irrigation to maintain optimal quality and yield at harvest. Recent droughts throughout the state coupled with strict groundwater regulations imposed through California's Sustainable Groundwater Management Act have created a major water scarcity problem for this lucrative industry, with nearly every almond producer in the state facing water supply challenges. Thus, there is a high demand for increased water use efficiency in almond production.

Regulated deficit irrigation (RDI), applying just the right amount of water or slightly less water at less sensitive growth stages, has been shown to greatly increase water use efficiency with only moderate reductions in quality and yield (DeJonge et al, 2015). However, proper implementation of RDI requires an accurate assessment of crop water status (DeJonge et al, 2015), which can be difficult to obtain. The most widely accepted method of determining a plant's water status is to measure its midday stem water potential (SWP) using a pressure chamber (Drechsler, 2019). Obtaining SWP measurements in this way is tedious and time consuming and thus is rarely used to assess crop water status in commercial settings. According to an Almond Board of California survey, less than 30% of almond growers use the pressure chamber to guide irrigation (Almond Board of California, 2019).

Numerous past studies have attempted to predict water status through other means, such as crop water stress indices or predictive models. In 1981 Jackson et al derived the widely known Crop Water Stress Index (CWSI), which is defined as the ratio of a crop's canopy temperature minus the air temperature, relative to the temperature differential of a well-watered plant and a non-transpiring plant

(DeJonge et al, 2015). While CWSI has been demonstrated to correlate well with SWP, it still requires calculation of a baseline using a well-watered crop leaf (often impractical to obtain) and its correlation with SWP has been found to vary widely throughout a growing season (Mueller et al, 2017). While the CWSI has its drawbacks, its consistent correlation with SWP indicates that its key parameters, namely air temperature, leaf/canopy temperature, relative humidity, and vapor pressure deficit, may be useful inputs in a predictive model of SWP.

In addition to crop water stress indices, remotely sensed vegetation indices (VI) may also be used to estimate crop water status. While Normalized Difference Vegetation Index (NDVI) is certainly the most widely used VI, Normalized Difference Red Edge Index (NDRE) has been demonstrated to correlate well with plant water content (Zhang 2019). By extension it may also be a useful parameter in the prediction of SWP.

In recent years, machine learning (ML) algorithms have emerged as a novel tool for modeling a wide variety of agricultural phenomena, including plant water status. Two common ML approaches are Artificial Neural Networks (ANN) and Random Forest (RF). ANNs are comprised of input, hidden, and output layers, each of which contain a number of neurons representing predictive variables. They are particularly useful for modeling of highly non-linear relationships (Virdnokar et al, 2020). A random forest model is made up of many decision "trees" which may be used for regression and classification. The defining attributes of this model type are the number of trees "n" and the number of variables used to make decisions at each tree, or "mtry". Past studies have determined n=500 as the optimal number of trees based on resulting error stabilization (Belgiu et al, 2016). "mtry" is typically set to the square root of the number of predictive variables (p), or p/3 if a more precise integer value is desired (Gislason et al, 2006). RF models are advantageous as they are highly resistant to overfitting compared with other model types (Virdnokar et al, 2020).

ML models have already been utilized in many cases for water stress prediction. In a 2018 study, an ANN model utilizing thermal indices along with canopy and air temperature data predicted SWP in

grapevines with relatively good accuracy [R^2=0.61] (Gutierrez et al, 2018). It was found that whether baseline temperatures for the indices were included as inputs or not, the model predicted SWP equally well. Several other studies have utilized remotely sensed multispectral imagery in ANN models to predict SWP with promising results, with coefficients of correlation ranging from 0.58 to 0.87 (Poblete et al, 2017; Romero et al, 2018). In another case, predictive performance of an optimum regression equation was compared with that of an ANN model, utilizing various soil and environmental parameters. The ANN model and regression equation resulted in coefficients of correlation between observed and predicted values of 0.93 and 0.85, respectively, indicating superiority of the ANN framework for water stress prediction. Applications of random forests for crop water status are less numerous, however Yang et al (2021) used a random forest model to predict canopy temperature of Chinese Brassica based on various meteorological data (R^2=0.77-0.9). Predicted temperatures were then used for calculation of CWSI.

Currently, applications of machine learning for water status prediction in almonds are relatively limited. In a 2018 study, an ANN model was used to predict water status in almonds with moderate success (R^2=0.78). In this case, the model was trained to predict $T_{diff\ dry}$, an alternative plant water status indicator, and utilized SWP as a predictive variable (Meyers et al, 2019).

Previous work has demonstrated that machine learning may be used effectively to predict water status in plants, based on soil and environmental variables, as well as those obtained through remote sensing. This study aims to investigate whether a machine learning based model can be trained to accurately predict SWP in almonds, utilizing a combination of soil, environmental, and remotely sensed variables. Additionally, it will investigate the ability of such a model to map SWP spatially across an entire orchard and will compare the performance of both ANN and RF based models.

Materials and Methods

Study Site and Experimental Layout

The site used in this study is a 1.6 hectare plot located at Nickels Soil Lab near Arbuckle, CA. The plot contains a total of 15 rows of 50 trees, with 5 rows each of Non-pareil, Butte, and Aldrich almond varieties. Throughout the two seasons during which the study took place, each row received a standard irrigation treatment determined by the grower. Each week irrigation was continually applied through driplines over two 24 hour periods.

For data collection, twenty focus trees were chosen throughout the orchard to serve as sampling locations, with four locations occurring in each non-pareil row (fig 1). At each location, a 1.5 m aluminum access tube was installed near the base of the tree for collecting soil moisture measurements using neutron attenuation with a neutron probe (model: Instrotek 503 Hydroprobe, San Francisco, CA). An additional ten locations were designated for validation measurements of SWP, for the purpose of assessing spatial SWP map predictions. Data collection was carried out once every one to two weeks, typically between the hours of 11am – 3pm.

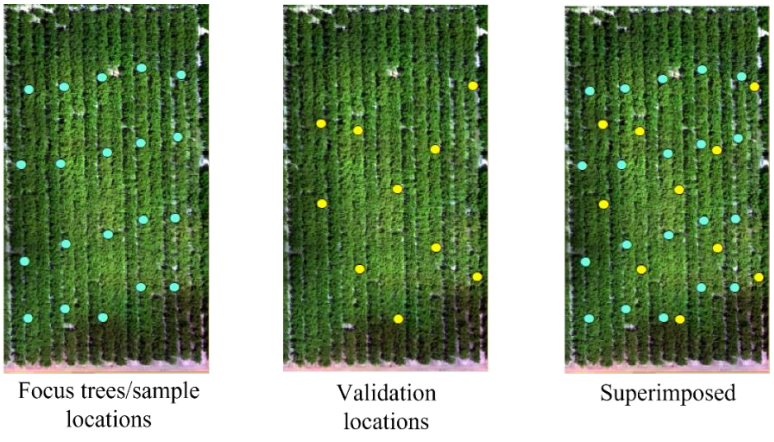


Fig 3) Locations of focus and validation trees for stem water potential measurements at the Nickels Soil Lab near Arbuckle, California.

Soil Moisture Measurements

On each measurement date, soil moisture was measured at each focus tree using a neutron probe. Five measurements were collected at each sampling location from 0.3 to 1.5m, in increments of 0.3m. Before beginning each day, a "standard count" was performed on the neutron probe to calibrate readings for current environmental conditions.

Midday Stem Water Potential Measurements

At each sample location, midday stem water potential was measured using a PMS Instrument Company Model 615 pressure bomb filled with nitrogen gas. To obtain midday stem water potential, a leaf is first covered for fifteen minutes using a Mylar foil bag, which removes the effect of solar radiation on the stem water potential reading. The leaf is then removed from the tree and inserted into a chamber within the pressure bomb, with the end of the severed stem protruding from the end of a small gasket. Then pressure is slowly added to the chamber using the nitrogen gas until water begins to bubble out of the end of the stem. The pressure at which this occurs is considered the tree's stem water potential (Dreschler, 2019). Prior to measuring, designated leaves were covered with small opaque bags in order to minimize the effect of solar radiation on the leaves' stem water potential. For each sample location, one leaf was measured from the focus tree, and one from the next tree immediately to the south. These two measurements were averaged and considered as the representative stem water potential for the location in question. Occasionally nitrogen gas would run low, in which case only one leaf would be measured at each location. This measurement was still considered representative of the water stress at its location.

For eight of the 2022 collection dates, ten extra measurements were obtained from each validation location for the purpose of spatial validation of SWP maps.

Figure 4) Left: PMS Model 615 Pressure Bomb used for SWP measurement. Right: Instrotek 503 Hydroprobe

Multispectral Imagery Collection

Multispectral imagery of the trees' canopies was collected on each measurement date using a DJI Matrice 100 drone equipped with a Micasense RedEdge multispectral camera. Flights were conducted at a height of 80m, utilizing an 85% front and side overlap for the images obtained. Directly before and after each flight, multispectral images of a reflectance calibration panel were captured in order to account for current light conditions.

Figure 5) Left to Right: DJI Matrice 100, Micasense Reflectance Calibration Panel, and Micasense RedEdge Multispectral Camera

Supplemental Environmental Data

Supplemental ET data for each date was obtained through California's CIMIS database to help improve the scalability of the research in any future applications. The data consisted of total daily ETo (mm/day), and was derived from the CIMIS station near Williams, CA, located approximately 16 km north of the study area. Daily ET values were considered representative of the entire orchard.

Processing of Multispectral Imagery

During the 2021 season, multispectral images were collected at too sparse of a scale, so creation of a complete orthomosaic of canopy reflectance values for each date was not possible. Instead, corresponding images for each sample location were selected based on coordinates embedded in their metadata, and canopy reflectance values were extracted using R Studio. Reflectance values were calibrated using a reflectance panel to account for specific ambient light conditions of each date.

For all 2022 collection dates, multispectral data was processed using Agisoft Metashape. By stitching together many overlapping images, an orthomosaic covering the entire area of the study site was produced. Using raster transformation functions within the software, each pixel was set to display its resultant NDRE value, based on the pixel's reflectance values for the red edge and near-infrared spectral bands. Raster files displaying NDRE were then exported to ArcMap, where they were all clipped to uniform boundaries and resampled to a pixel size of 0.2 meters to decrease model processing time. Using GPS coordinates of each sample location, NDRE values were extracted from each raster to match with other point data.

Artificial Neural Network Training and Sensitivity Analysis

Following collection and processing of all 2021 data, initial modeling and sensitivity analyses were carried out. MATLAB's Deep Learning Toolbox was used to set up and train artificial neural networks. Initially, eight neural network models were trained with different combinations of input parameters and compared based on their performance. The best model utilized soil moisture at 0.3-1.5m

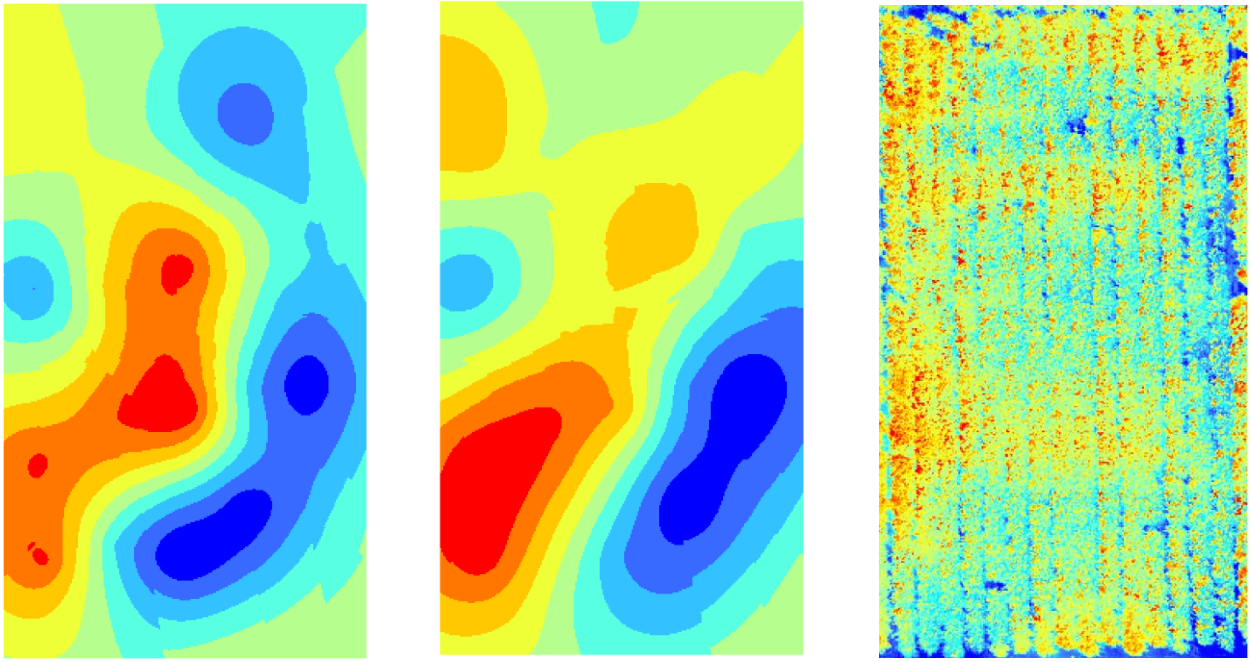
depths, ET, and NDRE as predictive variables. Basic sensitivity analysis was performed by retraining the model numerous times, with a different parameter excluded each time. The most important variable was determined to be soil moisture, followed by NDRE, and ET. Sensitivity analysis of individual soil moisture depths was not performed at this stage.

Following collection and processing of all data for the 2022 growing season, four more neural network models were trained. In order to simplify future model applications, soil moisture data for 1m through 1.5m depth was omitted. The first new ANN to be trained utilized soil moisture at 0.3-0.6m, daily ETo, and NDRE as inputs. The second utilized the same predictors but was trained to predict raw SWP values rather than baseline-adjusted SWP. The third included elevation data as an additional predictor, and the fourth utilized soil moisture at 0.3, 0.6, and 1.5m depths, ETo, and NDRE. This was performed to mirror the exact parameters used to train random forest models, and determine if this parameter combination also produced better results in an ANN model. Based on overall accuracy of prediction, the first and second models were both selected for generation of maps, in order to gain further insight into the effect of predicting baseline-adjusted versus non-baseline SWP values.

SWP Mapping using Artificial Neural Network Models

To generate maps of SWP using the selected ANN models, raster layers were first created for each input parameter using ArcGIS Desktop 10.8.1. Maps were only generated for dates on which extra validation SWP measurements had been collected, which were May 17, May 27, June 1, June 14, July 5, July 29, August 3, and August 18 (2022). To do so, kriging was performed using the coordinates and matching data of each sample location, using ArcGIS 10's default measurement variation of 100%. The kriging was fit to the boundaries of the previously generated NDRE layers and sampled to the same pixel size (fig 2). Each matching set of input layers was then exported to R Studio, where their pixel values were extracted and transformed into a singular array in which each column represented a single input parameter. The array was then exported to MATLAB, where it was fed into the ANN model, resulting in an output array of predicted SWP values. The output array was then exported to R studio, where it was

then transformed back into a raster layer with the same boundaries and resolution as before. Each map was then displayed in ArcMap, where predicted SWP values were extracted from all sample and validation locations for further analysis.



Soil Moisture (0.3m) Soil Moisture (0.6m) NDRE

Fig 6) Examples of kriging based spatial input layers used for SWP mapping. Left: Soil Moisture (0.3m), Center: Soil Moisture (0.6m), Right: Remotely sensed Normalized Difference Red Edge Index

Random Forest Forward Feature Selection

To train random forest models for SWP prediction, an R package called CAST was used which allows for spatial-temporal modeling using machine learning (Meyers, 2018).

Initially, a "forward feature selection" (FFS) was performed on the training dataset with all predictive variables included (fig 3). The analysis showed that by starting with two predictive variables and iteratively increasing the number of predictors, the R^2 value of the resulting model's predictions stopped improving after five predictors had been added. The FFS function reported soil moisture at 0.3 and 0.6 meters, NDRE, and ET to be the strongest predictive variables. Additionally, an initial random

forest model was trained with all input variables and a "variable importance" function was applied, which consistently ranked 1.5 meter soil moisture as the fifth most important variable (fig 4). Thus, the final combination of predictive variables for random forest modeling was selected as 0.3, 0.2, and 1.5 meter soil moisture, ET, and NDRE.

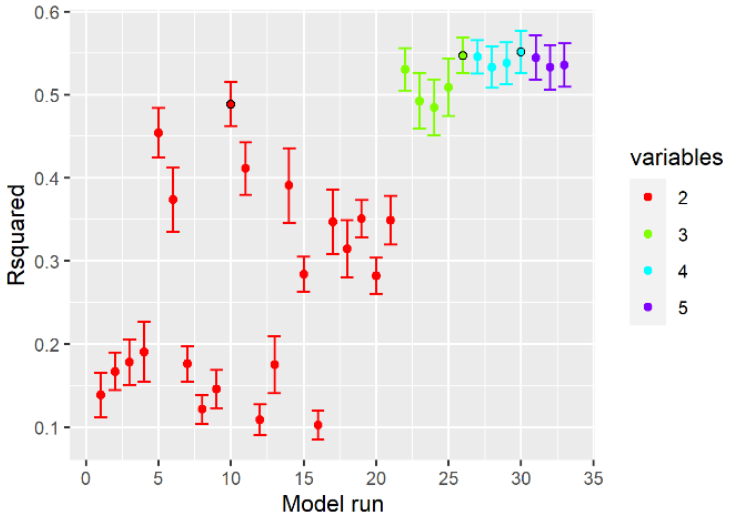


Fig 7) Results of Feed Forward Selection Analysis

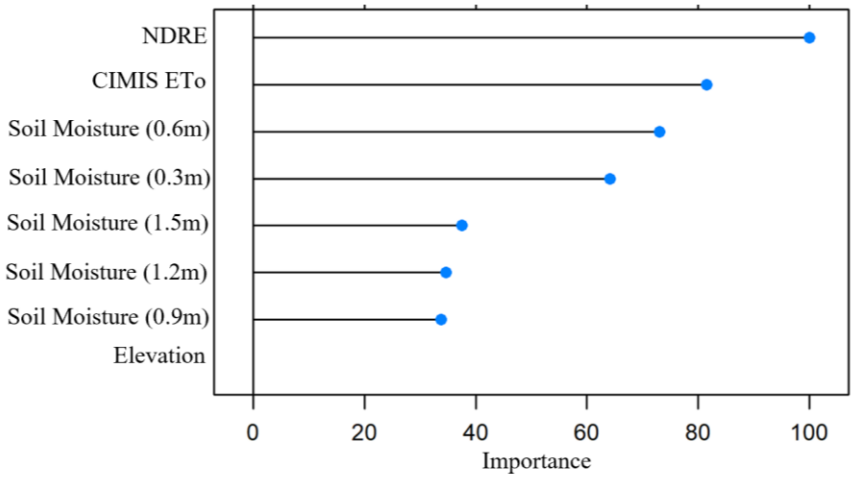


Fig 8) Results of Variable Importance Analysis with all Potential Predictors

Model Training and Target Oriented Validation

Using the designated input parameters, two random forest models were trained with 500 "trees" and an "mtry" value of two, based on the recommendation of using a value of p/3 in which p represents the number of predictive variables. A K-fold function with ten repetitions was used for model validation. One model was trained to predict baseline adjusted SWP while the other predicted non-baseline values.

In addition to regular model training and validation, a target-oriented validation was also performed. Rather than random folds, this method utilized specified folds in which the entire time series data of a different sampling location was omitted for each, using the CAST package's "CreateSpaceTimeFolds" function. This type of validation provided insight into the model's ability to predict stem water potential spatially. The same variable importance analysis function was also applied to the results of the target-oriented validation in order to understand the importance of each variable as it pertains to spatial prediction specifically.

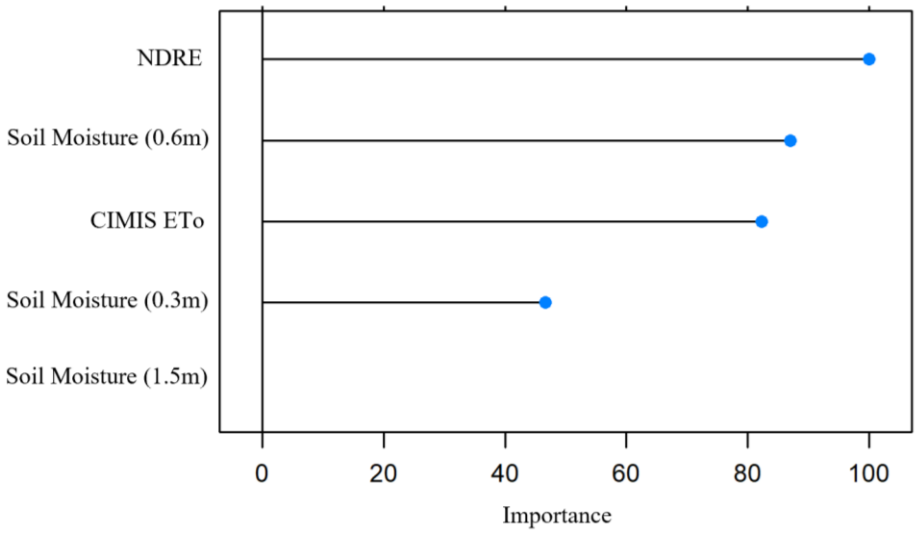


Figure 9) Results of Variable Importance Analysis on Target-Oriented Validation

SWP Mapping using Random Forest Model

Once sufficient random forest models had been trained and validated for both baseline adjusted and non-baseline SWP, maps were generated with the same input raster layers used for ANN map generation. For each date, the corresponding raster layers were imported to R Studio, where they were input to the model, which then returned a map of the predicted SWP.

Results

Results of Single Season ANN Modeling

Following the 2021 growing season, nine preliminary ANN models were trained with varying predictive variables. For all models, the coefficient of correlation between observed and predicted SWP ranged from 0.65 to 0.92. The best performing model utilized soil moisture, ET, and NDRE as predictive variables.

Final ANN Models

Following collection of all 2022 data, four additional ANN models were trained with two full seasons of data. For these models, coefficient of correlation between observed and predicted values ranged from 0.74 to 0.8 for training data, and from 0.75 to 0.88 for testing data. Compared with ANN 1, the normalized RMSE of ANN 2 increased by about 13 percent for both the training and testing datasets, indicating that training the ANN models to predict baseline adjusted SWP values may lead to stronger results. Model 3 included the same data as model 2, with the addition of elevation as a predictor. No significant increase in model performance was observed, indicating that the addition of elevation data did not increase predictive power. Model 4 was trained to reflect the performance of ANN modeling with the best combination of parameters from random forest modeling. All goodness-of-fit statistics showed a

decrease in model performance from model 2, indicating that the best parameter combination for random forest SWP modeling is not necessarily the best for ANN SWP modeling.

Table 2) Performative Statistics for all Models

	Training Data Performance				Testing data Performance			
Model	**R^2**	**rmse**	**nRMSE**	**IOA**	**R^2**	**rmse**	**nRMSE**	**IOA**
ANN 1	0.8	3.08	0.12	0.7	0.88	2.96	0.12	0.74
ANN 2	0.74	3.39	0.14	0.66	0.81	3.24	0.14	0.72
ANN 3	0.75	3.26	.14	0.67	0.79	3.84	0.16	0.71
ANN 4	0.75	3.34	0.13	0.67	0.75	3.56	0.16	0.67
RF 1	na	na	na	na	0.6	3.32	0.13	na
RF 2	na	na	na	na	0.57	3.36	0.14	na

**abbreviations: ANN: Artificial Neural Network, RF: Random Forest, IOA: Index of Agreement, rmse: Root mean squared error, nRMSE: Normalized Root Mean Squared Error*

Table 3) Descriptions of names, input parameters, and SWP prediction type for all models

Model Name	Parameters	SWP Prediction
ANN 1	SM (0.3-0.6m), ET, NDRE	Baseline Adjusted
ANN 2	SM (0.3-0.6m), ET, NDRE	Raw
ANN 3	SM (0.3-0.6m), ET, NDRE, Elevation	Raw
ANN 4	SM (0.3, 0.6, 1.5m), ET, NDRE	Raw
RF 1	SM (0.3, 0.6, 1.5m), ET, NDRE	Baseline Adjusted
RF 2	SM (0.3, 0.6, 1.5m), ET, NDRE	Raw

**abbreviations: ANN: Artificial Neural Network, RF: Random Forest, SM: Soil Moisture, ET: Evapotranspiration, NDRE: Normalized Difference Red Edge Index, SWP: Stem Water Potential*

ANN SWP Map Performance

Following generation of SWP maps for designated dates using the selected ANN models, predicted SWP values were extracted from all sampling locations as well as extra validation locations. These values were plotted against corresponding observed values, allowing for insight into the models' abilities to predict SWP spatially. For spatially predicted values at sampling locations, coefficient of correlation with observed values was much lower than for non-spatial point data, ranging from 0.2 to 0.3. RMSE and normalized RMSE both increased significantly as well. For extra validation locations, coefficient of correlation decreased slightly more, and RMSE and normalized RMSE both increased slightly. This makes sense, as model uncertainty would increase further the more you move away from data collection locations.

Random Forest Variable Importance Analysis

A variable importance analysis was performed in R Studio on a random forest model which used all potential predictors as inputs (soil moisture for 0.3 through 1.5 m, CIMIS ET, elevation, and NDRE). The analysis function was applied to both the original training data of the model as well as the results of a target-oriented validation which illustrated the model's ability to predict SWP spatially (fig 5). The analysis performed on the training dataset revealed 0.6 m soil moisture, CIMIS ETo, NDRE, and 1.5 m soil moisture to be the most important predictors, in that order. 0.3 m , 0.9 m, and 1.2 m soil moisture were reported to have little to no significance in regard to the model's prediction, and elevation was reported to have no significance whatsoever. The analysis performed on the target-oriented validation produced nearly the same result, with the exception of 0.3 m soil moisture being listed as more important than 1.5 m moisture. This was interpreted as an indication that while 0.3 m soil moisture may not be significant in prediction based solely on point data, it becomes more useful when the model is extrapolated spatially between point data locations.

Random Forest Feed Forward Selection

A feed forward selection function was applied to all predictive data, the results of which can be viewed in figure 3. The analysis showed that as models were successively trained with more predictive variables, the coefficient of correlation for the resulting model stopped improving after a fifth parameter was added. This finding is in agreement with the variable importance analysis, which indicated five parameters that were significant to model prediction.

Random Forest Modeling Results and Map Performance

For the baseline-adjusted and non-baseline random forest models, coefficients of correlation were 0.57 and 0.6, respectively, and nRMSE values were 0.14 and 0.13, respectively. These results indicate an increase in predictive performance compared with ANN models. Due to the nature of the R Studio package used, index of agreement was not possible to estimate for random forest models.

A slight increase in performance was observed when comparing baseline-adjusted with non-baseline model results. A roughly 5 percent increase in coefficient of correlation was observed, while nRMSE decreased by approximately 0.7 percent.

Performance of SWP maps generated by random forest models generally increased compared with that of ANN. For non-baseline SWP maps, coefficient of correlation at validation locations was 0.4, while nRMSE and index of agreement were 0.19 and 0.56, respectively. Contrary to model training results, performance of baseline adjusted SWP maps was lower by comparison (R^2=0.17, nRMSE=0.24, IOA=0.52). The best spatial mapping results were produced by non-baseline adjusted random forest maps, while the best model training results occurred with baseline-adjusted random forest models. The reason behind this contradiction is not known. However it is possible that the single hourly values of relative humidity and air temperature used in baseline calculation do not account for true variation throughout the orchard, leading to a slight increase in model uncertainty.

Comparison of RF and ANN Models

No significant difference was observed in model performance for ANN models versus RF. RF based maps produced noticeably more accurate predictions at data sampling locations than ANN based maps, but prediction accuracy at validation measurement locations was essentially the same for both map types.

Table 4) Summary of Performance Statistics for all Maps

	Sampling Locations				Validation Locations			
Model	**R^2**	**rmse**	**nRMSE**	**IOA**	**R^2**	**rmse**	**nRMSE**	**IOA**
ANN 1	0.3	3.06	0.21	0.52	0.29	3.64	0.23	0.54
ANN 2	0.2	3.11	0.2	0.46	0.18	3.81	0.2	0.5
RF 1	0.6	2.74	0.19	0.59	0.17	3.84	0.24	0.52
RF 2	0.47	2.71	0.18	0.56	0.4	3.6	0.19	0.56

**abbreviations: ANN: Artificial Neural Network, RF: Random Forest, IOA: Index of Agreement, rmse: Root mean squared error, nRMSE: Normalized Root Mean Squared Error*

Comparison of Model and Map Performance

For all models, a significant decrease in performance was observed when comparing respective map performance to model performance. Generally, coefficients of correlation decreased by roughly fifty percent, and nRMSE values roughly doubled. In models for which index of agreement could be calculated, an approximate twenty five percent decrease in value was observed.

For RF models, map performance at sampling locations remained mostly the same as model performance, while performance at validation locations was noticeably worse. For ANN models, map performance at sampling locations was significantly worse than model performance, but remained mostly constant when considering performance at validation points. Thus, no significant difference in map

performance at validation locations was observed between ANN and RF based maps, but RF based maps performed significantly better than ANN maps at sampling locations.

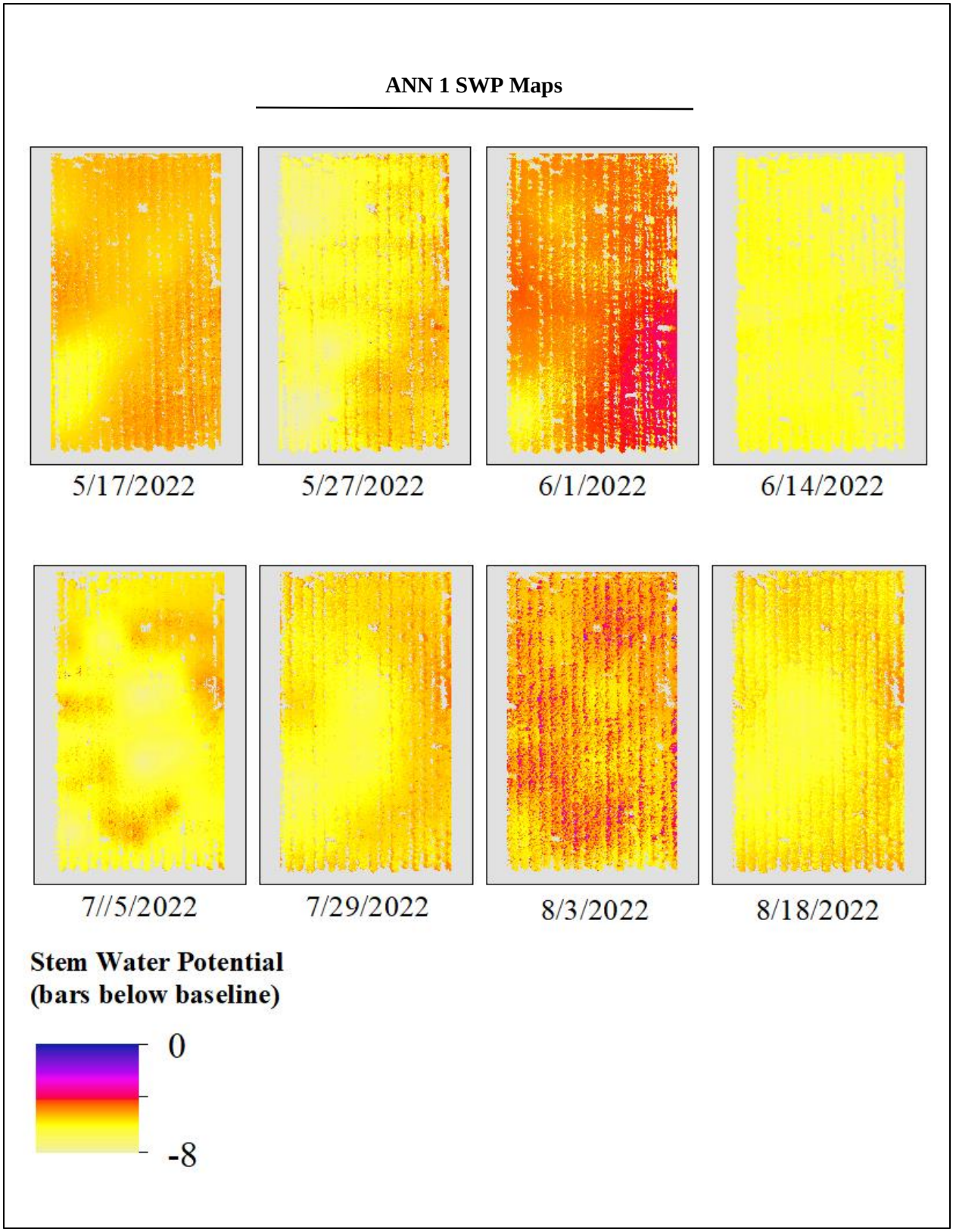


Fig 10) ANN baseline SWP Maps

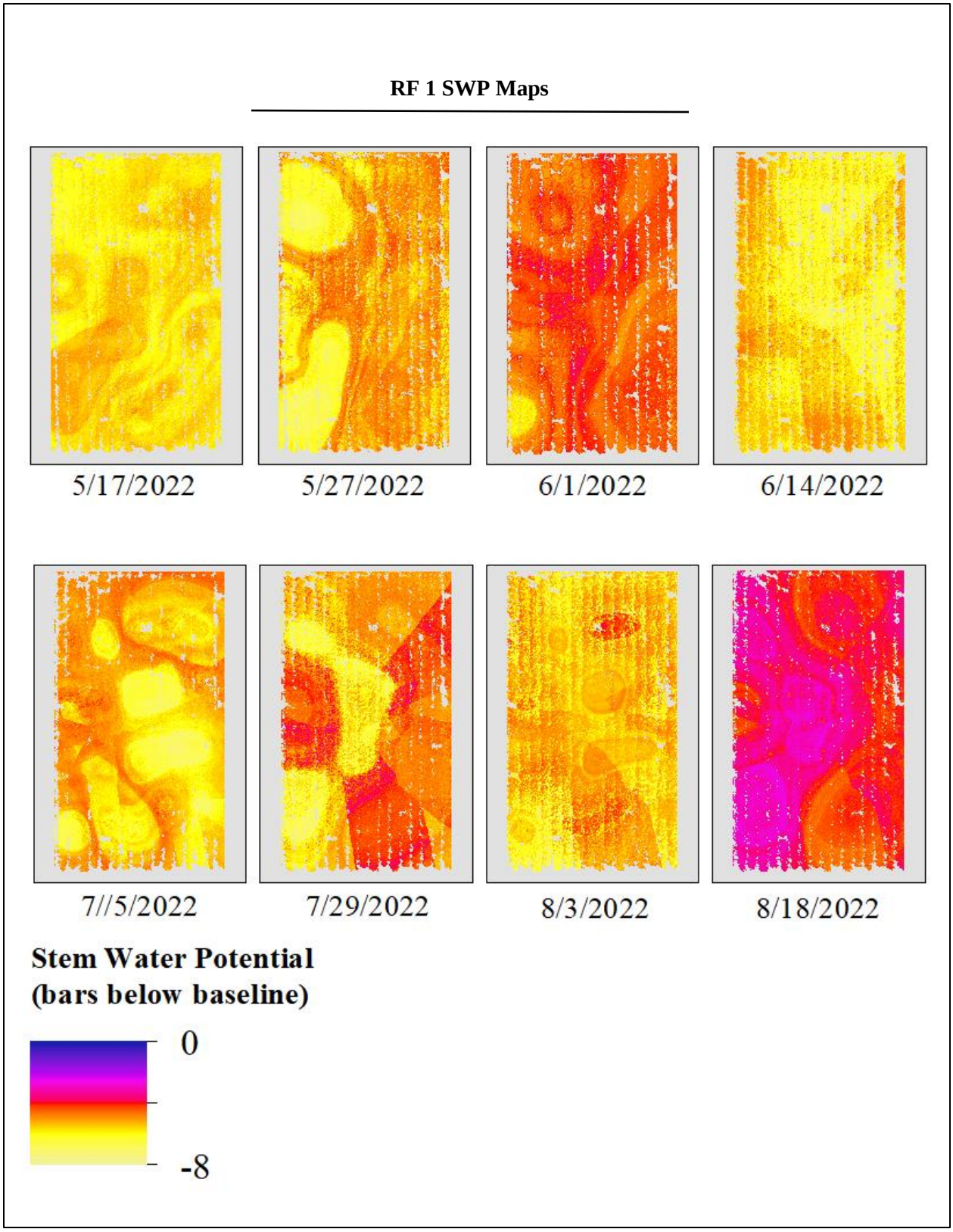


Fig 11) RF Baseline SWP Maps

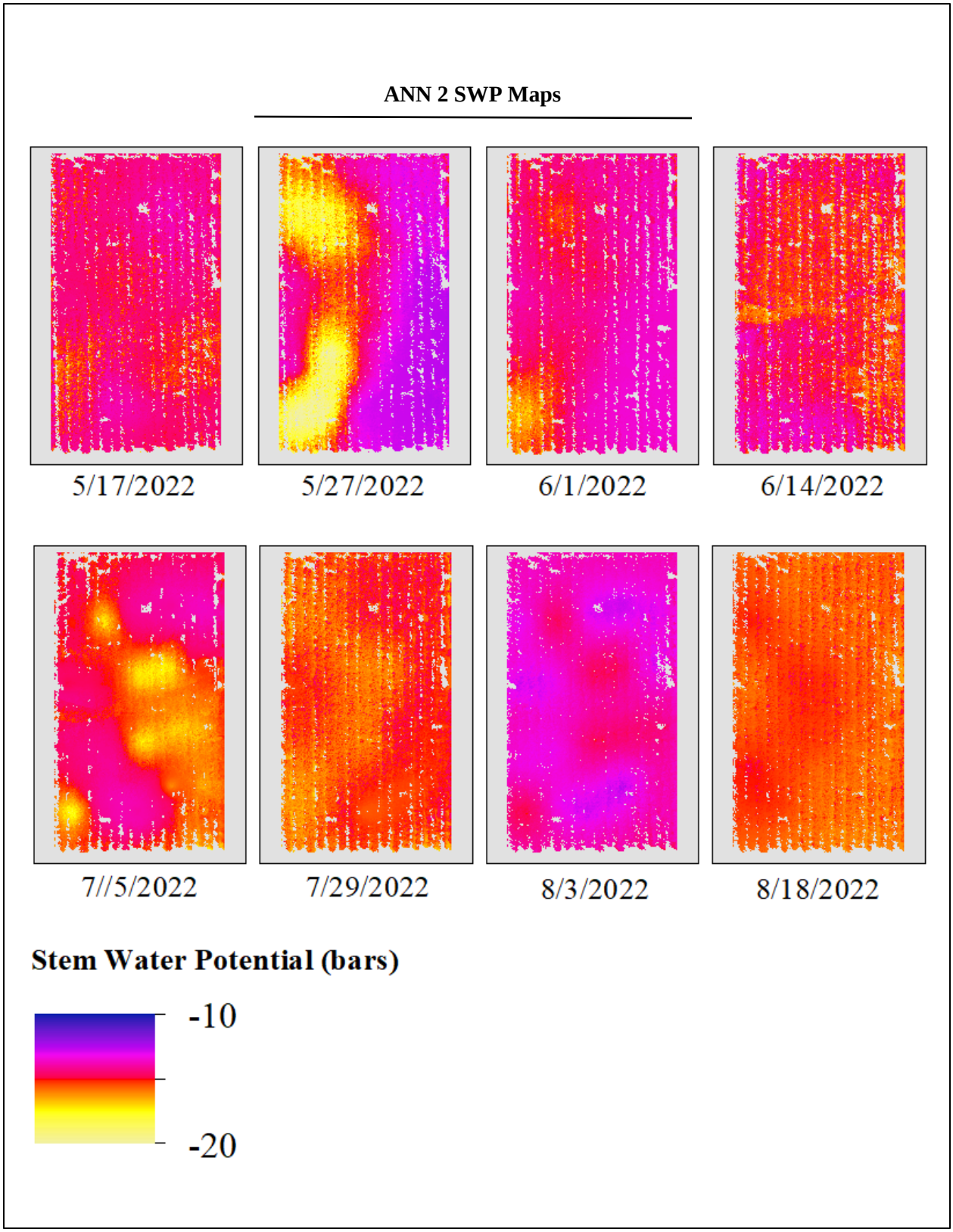


Fig 12) ANN Raw SWP Maps

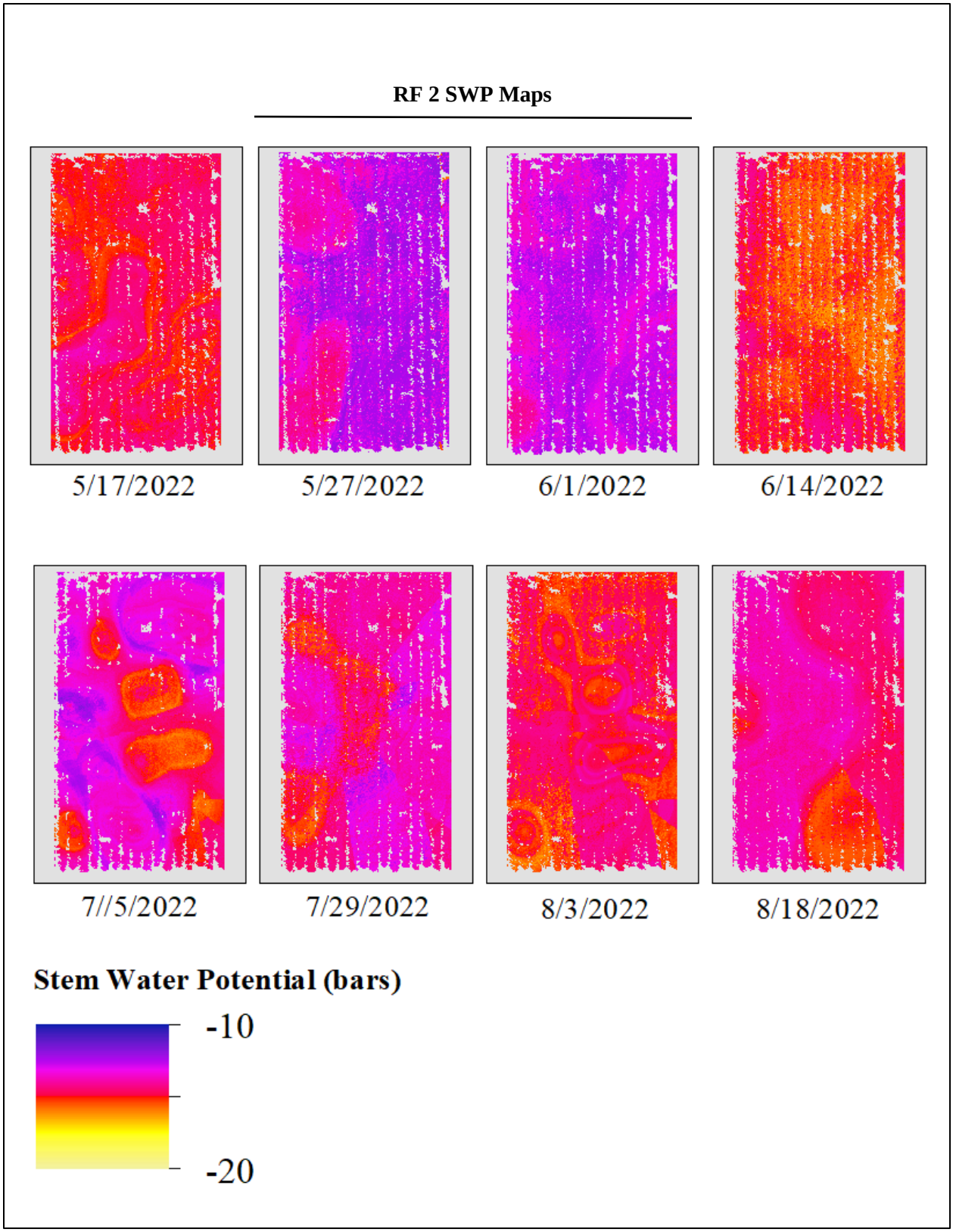


Fig 13) RF Raw SWP Maps

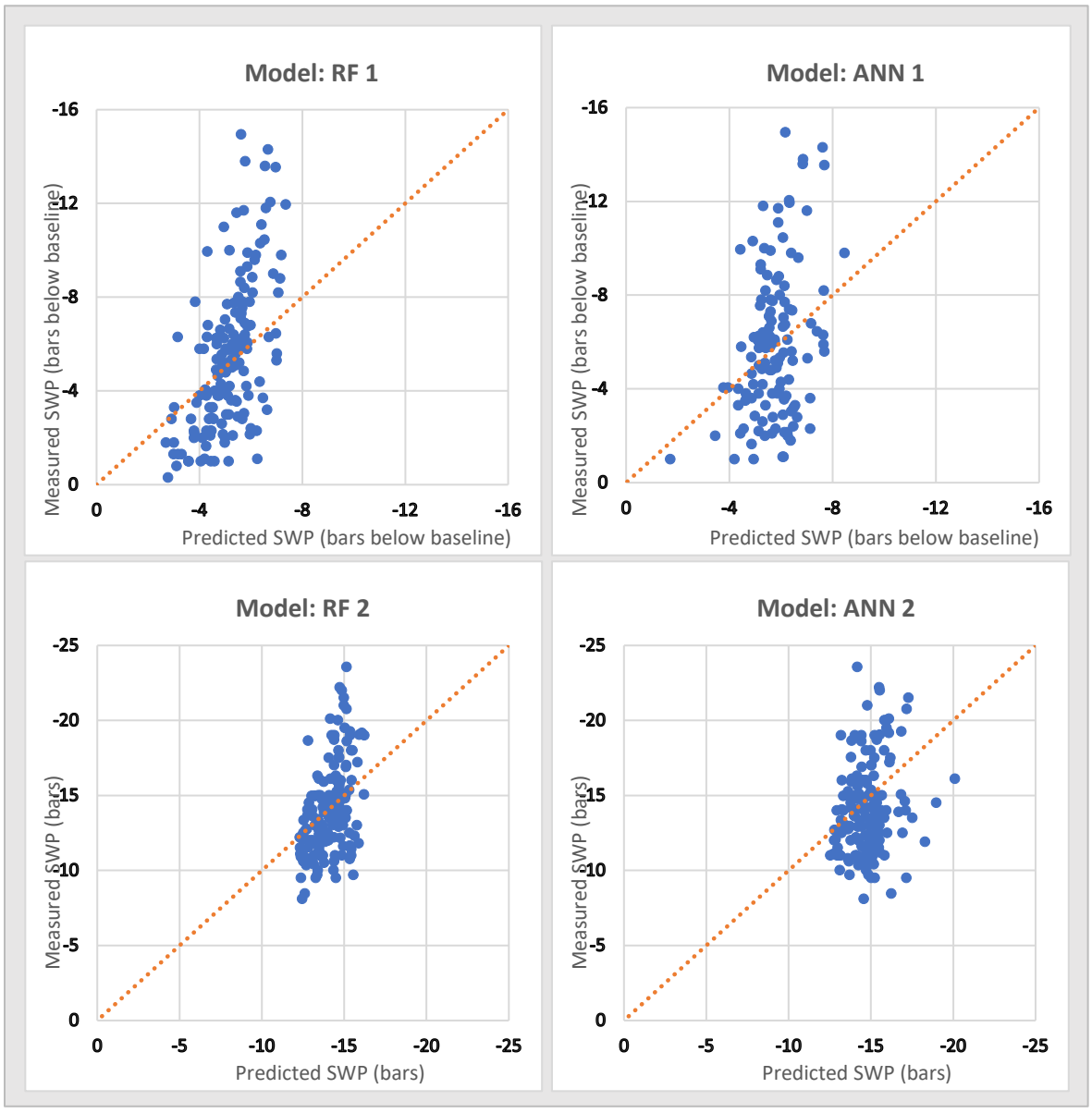


Fig 14) Map performance at data sampling locations

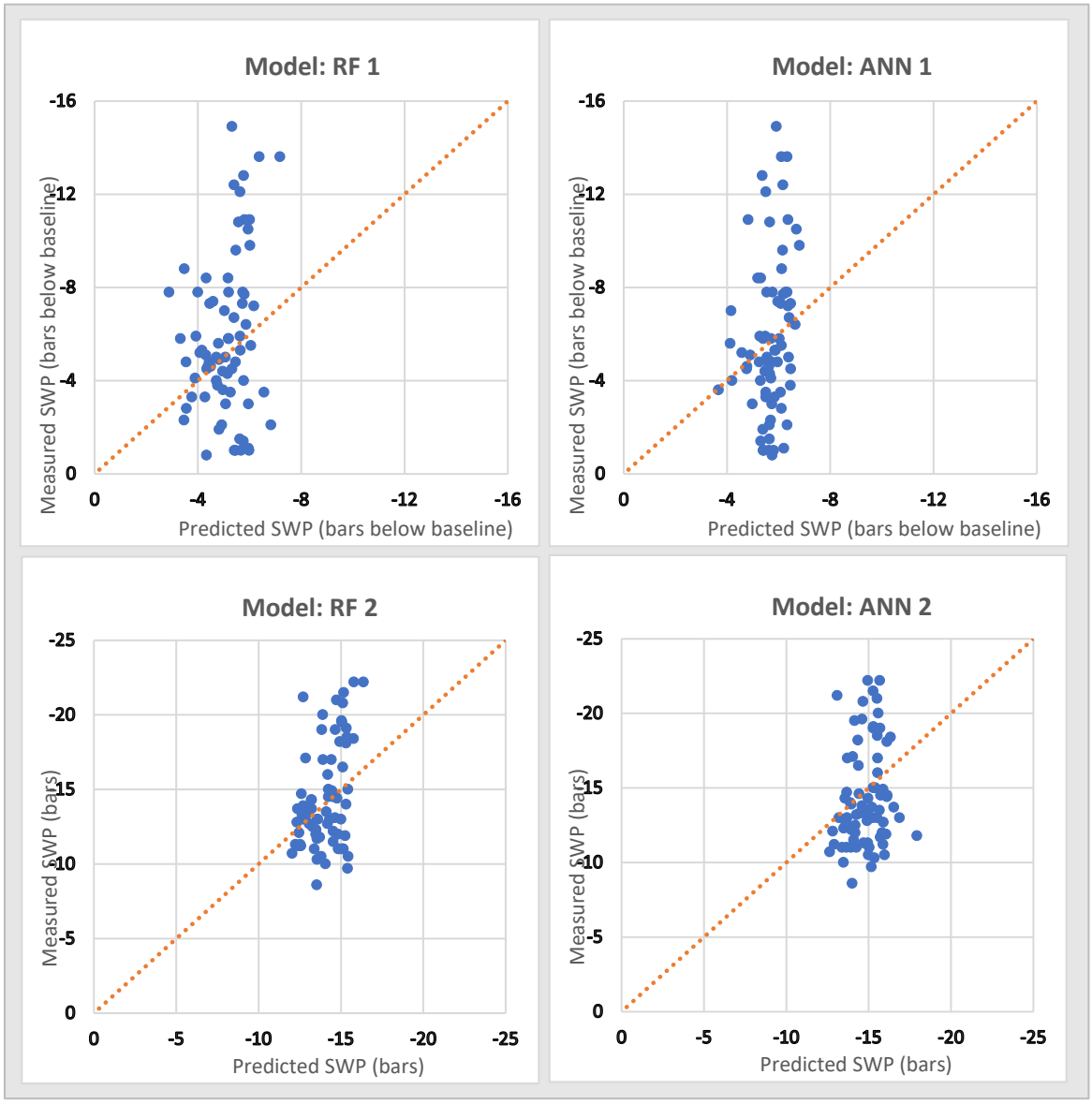


Fig 15) Map performance at extra validation locations

Discussion

Results of the experiment have demonstrated the ability of machine learning models to predict spatial SWP in almonds. Across all models, an average coefficient of correlation of R^2=0.73 and average normalized RMSE of 0.14 were observed between measured and predicted values, which is largely in agreement with the findings of recent similar studies. Poblete et al (2017) and Pocas et al (2017) used ML frameworks to predict grapevine SWP with R^2=0.58-0.87 and R^2=0.77, respectively. Meyers et al (2019) predicted $T_{diff\ dry}$, (novel plant water status indicator) with R^2=0.78 (average). Based on sensitivity analysis from both RF and ANN models, it seems that spectral reflectance data was the strongest predictor of water stress, followed closely by root zone soil moisture. This confirms the findings of Zhang et al (2019) that Normalized Difference Red Edge Index can vary significantly in response to changes in canopy water content. The same study also found the red edge spectral band (integral to NDRE calculation) to correlate well with canopy water content (R^2=0.78). Romero et al (2018) used an ANN model to predict vineyard SWP using only remotely sensed vegetation indices (R^2=0.72). Both the literature and this study's results indicate the potential for development of future almond SWP models based entirely on remotely sensed data, eliminating the need for in-situ sensors and labor. Going forward, this could greatly increase the commercial viability of similar workflows for assessment of water stress.

In both RF and ANN models, map prediction accuracy was significantly less accurate than that of the base models themselves. It seems that a higher spatial sampling density for input variables is necessary in order to account for the true spatial variation found throughout an orchard. Additionally, this performance gap may be addressed with the inclusion of more remotely sensed parameters, such as leaf canopy temperature, which can be precisely measured on a pixel-by-pixel scale.

When comparing baseline-adjusted SWP predictions to raw SWP predictions, little difference was observed in model performance with point data. This agrees with results published by Gutierrez et al (2018), who used thermal indices as predictors in ANN models to predict grapevine SWP. It was found

that whether or not baseline temperatures were included in thermal indices, the resulting models performed equally well. However, raw SWP showed a slight increase in accuracy for spatial map predictions. As previously mentioned, the single hourly value parameters used in baseline calculation likely do not account for variations occurring within the orchard. Furthermore, baseline calculations are by nature an approximation, which introduces further uncertainty to SWP models. It is possible that training models with larger datasets could reveal these uncertainties to a greater degree.

Both RF and ANN based models performed very similarly, however RF based maps showed more accurate predictions at data sampling locations. In ANN based maps, prediction accuracy roughly the same at both data sampling and validation locations. At sampling locations, the RF framework seemed to maintain a degree of consistency between model and map predictions that the ANN framework did not. Additionally, as RF models are highly resistant to overfitting, it seems reasonable to suggest that the RF modeling framework is better suited to water stress prediction for almond trees.

CHAPTER 3: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Conclusions

1. Overall results indicate that both ANN and RF based models can reliably predict SWP in almond orchards. While model performance was essentially the same for both RF and ANN models, it seems reasonable to conclude that RF models are better suited to SWP prediction in almonds, based on the superior performance of RF based maps at data sampling locations. However, it is possible that this gap in performance could be removed through the use of larger training datasets for both model types.

2. As no significant difference in model performance was observed when baseline adjustments were included in target SWP values, it may be concluded that training models to predict "raw" SWP values is a better approach. Baseline calculation for midday SWP is an approximation at best and is calculated based on air temperature and relative humidity. Thus, the noisiness of datasets is bound to increase with the inclusion of baselines. Additionally cross-correlation may occur between SWP baseline and ET data calculations, adding redundancy to models. However, it is worth noting that modeling with larger datasets may reveal more noticeable differences in performance.

3. None of the modeling approaches detailed in this study can properly account for the true spatial variability in SWP, based on the consistent decreased performance observed for validation measurement locations.

Recommendations for Future Research

1. The results of this study are limited by the size of training datasets used. The laborious nature of SWP measurement, multispectral drone mapping, and soil moisture measurement make large datasets difficult to obtain. Future studies should utilize more training data whenever possible. This may help to eliminate some of the ambiguities encountered while analyzing model performance, and likely would increase model and map performance in all aspects.

2. In order to increase spatial map performance, future studies should increase spatial sampling density if possible. Additionally, the inclusion of more "high spatial density" parameters, such as remotely sensed canopy temperature may help to improve performance. Alternate spatial interpolation methods may also be investigated in order to improve spatial representation of predictive variables.

3. Future research should investigate the use of SWP maps as an irrigation decision guide, in which irrigation events are triggered once predicted SWP reaches a high enough threshold. This may be compared with a simple irrigation plan, such as the one applied to the study plot for the 2021 and 2022 growing seasons. While frequency and timing of data collection in this study was limited by logistics and scheduling, future work should aim to consistently collect data directly before or after irrigation events. Studies should investigate the effect of SWP maps as irrigation guides on water usage, almond yield, and almond quality.

References

Almond Board of California. (2019). Supplemental File 1 Irrigation.

Belgiu, M., & Druaguct, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.

Berni, J., Zarco-Tejada, P. J., Suarez, L., & Fereres, E. (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. IEEE Transactions on Geoscience and Remote Sensing, 47(3), 722–738. https://doi.org/10.1109/tgrs.2008.2010457

Campbell, G.S., Campbell, M.D., 1982. Irrigation scheduling using soil moisture measurements: theory and practice. In: Hillel, D.J. (Ed.), Advances in Irrigation, vol. 1. Academic Press, New York, pp. 25–42.

Carter, G.A. (1991) Primary and secondary effects of water content on the spectral reflectance of leaves. American Journal of Botany, vol. 78, no. 7, pp. 916– 924.

Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 2014, 92, 79–97.

David Goldhamer & Robert Beede (2004) Regulated deficit irrigation effects on yield, nut quality and water-use efficiency of mature pistachio trees, The Journal of Horticultural Science and Biotechnology, 79:4, 538-545, DOI: 10.1080/14620316.2004.11511802

DeJonge, K. C., Taghvaeian, S., Trout, T. J., & Comas, L. H. (2015). Comparison of canopy temperature-based water stress indices for maize. Agric. Water Mgmt., 156, 51-62. https://doi.org/10.1016/j.agwat.2015.03.023

Dhillon, R., Rojo, F., Upadhyaya, S. K., Roach, J., Coates, R., & Delwiche, M. (2018). Prediction of plant water status in almond and walnut trees using a continuous leaf monitoring system. Precision Agriculture, 20(4), 723-745. doi:10.1007/s11119-018-9607-0

Drechsler, K., Kisekka, I., & Upadhyaya, S. (2019). A comprehensive stress indicator for evaluating plant water status in almond trees. *Agricultural Water Management*, *216*, 214–223. https://doi.org/10.1016/j.agwat.2019.02.003

Durigon, A., & Lier, Q. D. (2013). Canopy temperature versus soil water pressure head for the prediction of crop water stress. Agricultural Water Management, 127, 1-6. doi:10.1016/j.agwat.2013.05.014

Espinoza, C.Z.; Khot, L.R.; Sankaran, S.; Jacoby, P.W. High Resolution Multispectral and Thermal Remote Sensing-Based Water Stress Assessment in Subsurface Irrigated Grapevines. Remote Sens. 2017, 9, 961.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. Pattern Recognition Letters, 27(4), 294–300.

Goldhamer, D. A. (2005). Tree water requirements and regulated deficit irrigation. In L. Ferguson (Ed.), Pistachio production manual (4th ed., pp. 103–116). Davis: Fruit and Nut Research and Information Center, University of California.

Gutierrez S, Diago MP, Fernandez- Novales J, Tardaguila J (2018) Vineyard water status assessment using on-the-go thermal imaging and machine learning. PLoS ONE 13(2): e0192037. https://doi.org/10.1371/journal. pone.0192037

Hsu, K.L.; Gupta, H.V.; Sorooshian, S. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res. 1995, 31, 2517–2530.

Idso, S.B., Jackson, R.D., Reginato, R.J., 1977. Remote-sensing of crop yields. Science 196, 19–25.

Idso, S., Jackson, R., Pinter, P., Reginato, R., & Hatfeld, J. (1981). Normalizing the stress-degree-day parameter for environmental variability. Agricultural Meteorology, 24, 45–55.

Jackson, R. D., Idso, S. B., Reginato, R. J., & Pinter, P. J. (1981). Canopy temperature as a crop water stress indicator. Water Resources Research, 17(4), 1133–1138. https://doi.org/10.1029/wr017i004p01133

Jones HG. Use of infrared thermometry for estimation of stomatal conductance as a possible aid to irrigation scheduling. Agricultural and forest meteorology. 1999; 95(3):139–149. https://doi.org/10.1016/S0168-1923(99)00030-1

King, B.; Shellie, K. Evaluation of neural network modeling to predict non-water-stressed leaf temperature in wine grape for calculation of crop water stress index. Agric. Water Manag. 2016, 167, 38–52.

Loggenberg, K., Strever, A., Greyling, B., & Poona, N. (2018). Modelling water stress in a shiraz vineyard using hyperspectral imaging and Machine Learning. *Remote Sensing*, *10*(2), 202. https://doi.org/10.3390/rs10020202

Martí, P., Gasque, M., & González-Altozano, P. (2013). An artificial neural network approach to the estimation of stem water potential from frequency domain reflectometry soil moisture measurements and meteorological data. Computers and Electronics in Agriculture, 91, 75-86. doi:10.1016/j.compag.2012.12.001

Meyers, H (2018). CAST [source code] https://cran.r-project.org/web/packages/CAST/index.html

Meyers, J. N., Meyers, J. N., Kisekka, I., Upadhyaya, S. K., Michelon, G. K., Kisekka, I., . . . Michelon, G. K. (2019). Development of an artificial neural network approach for predicting plant water status in almonds. Transactions of the ASABE, 62(1), 19-32. doi:10.13031/trans.12970

Möller, M.; Alchanatis, V.; Cohen, Y.; Meron, M.; Tsipris, J.; Naor, A.; Ostrovsky, V.; Sprintsin, M.; Cohen, S. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. J. Exp. Bot. 2007, 58, 827–838.

Naor, A. 2008. Water stress assessment for irrigation scheduling of deciduous trees. Acta Horticulturae 792.

Parker, T. A., Palkovic, A., & Gepts, P. (2020). Determining the genetic control of common bean early-growth rate using unmanned aerial vehicles. *Remote Sensing*, *12*(11), 1748.

Poblete, T., Ortega-Farías, S., Moreno, M., & Bardeen, M. (2017). Artificial Neural Network to Predict Vine Water Status Spatial Variability Using Multispectral Information Obtained from an Unmanned Aerial Vehicle (UAV). Sensors, 17(11), 2488. doi:10.3390/s17112488

Pôças, I., Gonçalves, J., Costa, P. M., Gonçalves, I., Pereira, L. S., & Cunha, M. (2017). Hyperspectral-based predictive modelling of grapevine water status in the Portuguese douro wine region. International Journal of Applied Earth Observation and Geoinformation, 58, 177–190. https://doi.org/10.1016/j.jag.2017.02.013

Romero, M., Luo, Y., Su, B., & Fuentes, S. (2018). Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management. Computers and Electronics in Agriculture, 147, 109-117. doi:10.1016/j.compag.2018.02.013

Starr, G. C. (2005). Assessing temporal stability and spatial variability of soil water patterns with implications for precision water management. Agricultural Water Management, 72(3), 223–243.

Sumner, D., William Matthews, Medellin-Azuara, J., & Bradley, A. (2014). (rep.). *The Economic Impacts of the California Almond Industry*.

Virnodkar, S., Pachghare, V., Patil, V. C., Jha, S. (2020). Remote sensing and machine learning for crop water stress determination in various crops: A critical review. Precision Agriculture, (21), 1121-1155. doi:https://doi.org/10.1007/s11119-020-09711-9

Yang, M., Gao, P., Zhou, P., Xie, J., Sun, D., Han, X., & Wang, W. (2021). Simulating canopy temperature using a random forest model to calculate the crop water stress index of Chinese brassica. *Agronomy*, *11*(11), 2244. https://doi.org/10.3390/agronomy11112244

Zhang, F., & Zhou, G. (2019). Estimation of vegetation water content using hyperspectral vegetation indices: A comparison of crop water indicators in response to water stress treatments for summer maize. *BMC Ecology*, *19*(1). https://doi.org/10.1186/s12898-019-0233-0