ESTIMATIVA DE UMIDADE DO SOLO POR MEIO DE APRENDIZADO DE MÁQUINA USANDO IMAGENS DE VEICULO AÉREO NÃO TRIPULADO (VANT)

ANDERSON LUIZ DOS SANTOS SAFRE[1]; CAIO NASCIMENTO FERNANDES[2]; JOÃO CARLOS CURY SAAD[3]

[1]*Aluno de Doutorado em Irrigação e Drenagem, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agronômicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: andersonsafre@gmail.com*
[2]*Aluno de Mestrado em Irrigação e Drenagem, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agronômicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: caionfernandes@hotmail.com*
[3]*Professor Titular, Departamento de Engenharia Rural, UNESP-Faculdade de Ciências Agronômicas, R. José Barbosa de Barros, 1780, CEP 18610-034, Botucatu-SP, Brasil. E-mail: joaosaad@fca.unesp.br*

1 RESUMO

A umidade do solo é um parâmetro importante para o cálculo da lâmina e manejo da irrigação, pois está diretamente relacionada ao conteúdo de água no solo. Técnicas de sensoriamento remoto aliadas a modelos estatísticos podem ser usadas para estimar a variabilidade espacial da umidade do solo, extrapolando medidas pontuais. O objetivo desse estudo foi determinar a umidade do solo por meio de algoritmos de *machine learning* (aprendizado de máquina) como *Support Vector Regression* (SVR), *Random Forests* (RF) e *Artificial Neural Networks* (ANN). Utilizou-se imagens multiespectrais de alta resolução adquiridas por meio de um Veículo Aéreo Não Tripulado (VANT) em uma área de feijão irrigado na Fazenda Experimental Lageado da Unesp, em Botucatu, SP, Brasil. Adotou-se como dados de entrada nos modelos, as refletâncias nas bandas do verde, vermelho, infravermelho próximo e o NDVI. Todos os algoritmos tiveram performance adequada, porém o modelo que melhor estimou a umidade do solo foi o SVR, com erro médio quadrático (RMSE) de 0,46 vol. % e coeficiente de determinação (R^2) de 0,71.

Palavras-chave: umidade do solo, aprendizado de máquinas, VANT, redes neurais.

SAFRE, A. L. S.; FERNANDES, C. N.; SAAD, J. C. C.
SOIL MOISTURE ESTIMATION THROUGH MACHINE LEARNING USING UNMANNED AERIAL VEHICLE (UAV) IMAGES

2 ABSTRACT

The soil moisture is an important parameter for the calculation of water depth and irrigation management since it is directly related to the soil water content. Remote sensing techniques combined with statistical models can be used to estimate the spatial variability of soil moisture, extrapolating point measurements. The objective of this study was to determine the soil moisture through machine learning algorithms such as Support Vector Regression (SVR), Random Forests (RF), and Artificial Neural Networks (ANN). High resolution multispectral images obtained by an Unmanned Aerial Vehicle (UAV) in an irrigated bean area at the

Experimental Lageado Farm at Unesp in Botucatu, SP, Brazil, were used. The reflectances in the Green, Red and Near Infrared bands along with the NDVI vegetation index were used as inputs for the models. All the algorithms performed well; however, the model that best fitted the data was the SVR, with mean square error (RMSE) of 0.46% of the estimated soil moisture and determination coefficient (R^2) of 0.71.

Keywords: soil moisture, machine learning, UAV, artificial neural networks.

3 INTRODUCTION

The availability of water and energy resources for agriculture is increasingly limited due to climate change and the growing demand for food. Agricultural systems, especially irrigated systems, require techniques and management that provide high levels of efficiency to preserve natural resources while simultaneously developing more competitive agricultural production systems. Therefore, soil moisture monitoring is one of the fundamental practices for efficient and rational management of irrigated agricultural systems, as it allows for better crop development and optimized use of both water and energy resources (HUISMAN et al., 2003; DUKES; ZOTARELLI; MORGAN, 2010; MONTESANO et al., 2015; BRITO et al., 2009; FREITAS et al., 2012; BRITO et al., 2014).

Soil moisture can be measured via direct or indirect methods. Moisture measurement via the direct method is determined by calculating the difference between the mass of the soil sample in its initial state and after drying. Soil resistance to electric current, neutron probes, capacitive sensors, and soil water tension are examples of indirect methods for determining soil moisture (TEIXEIRA; COELHO, 2005; DOBRIYAL et al., 2012). Remote sensing is also an indirect method, as images captured by cameras attached to satellites or unmanned aerial vehicles (UAVs) allow for the correlation of soil moisture with variations in reflected electromagnetic radiation via statistical methods.

Measuring soil water tension with equipment called tensiometers is one of the best-known techniques for indirectly determining moisture. In addition to being less expensive than other techniques, it offers easy handling, high accuracy, and the possibility of automating the reading system (ARRUDA et al., 2017) and the irrigation system itself (MONTESANO et al., 2015). Accurately quantifying the soil water content is necessary to aid producers in their decision-making regarding when and how much to irrigate, thus ensuring efficient and rational management (THALHEIMER, 2013) and providing the appropriate amount of water for crop development, resulting in productivity gains.

In the initial development of the tensiometer, the chemical element mercury was used as a measuring scale, implemented by Livingston in 1908 (OR, 2001), which demonstrated the ability to measure the soil matric potential (LIBARDI, 2005). The soil in contact with the porous capsule of the tensiometer reached equilibrium. As the soil moisture decreases, the water present in the tensiometer is absorbed by the generated matric tension, reducing the internal pressure of the system and causing an increase in tension in the tensiometer. The tension reading is compared with the soil's characteristic water retention curve, and thus, the soil moisture can be determined (CAMARGO; GROHMANN; CAMARGO, 1982).

Precision agriculture allows producers to monitor the specific conditions

of each location, enabling extremely efficient management, precisely because its main principle is the use of farm-specific parameters. In precision agriculture, spatial information systems, such as geographic information systems (GISs), remote sensing tools, and global positioning systems, can be highlighted.

Remote sensing can be defined as the method of acquiring information about a specific phenomenon or behavior on the Earth's surface without physical contact (JENSEN, 2009). This method allows the identification of specific target characteristics through interaction with electromagnetic radiation (EMR) (ROSA, 2009). For this purpose, sensors attached to satellites, aircraft, UAVs, and other platforms are used.

The EMR is energy that travels at the speed of light, whether in the form of particles or electromagnetic waves, and does not require a physical medium for propagation (ROSA, 2009). The EMR comprises the electromagnetic spectrum, which encompasses all wavelengths, from gamma rays to radio waves (NOVO, 2008). When propagating through space, the flux of electromagnetic radiation may or may not interact with objects or surfaces and thus be reflected, absorbed, or transmitted (ROSA, 2009; PONZONI; SHIMABUKURO, 2010). Sensors used in remote sensing record target reflectance, which is highly important in agriculture, as it can extract important information about the metabolic state of crops.

Recently, nonlinear statistical methods, such as machine learning, have been applied to remote sensing data to estimate physical parameters. Algorithms such as *random forest* (BREIMAN, 2001), *support vector machines* (KOVAČEVIC; BAJAT; GAJIC, 2010; PRIORI; BIANCONI; CONSTANTINI; 2014), artificial neural networks (AITKENHEAD et al., 2013; SILVEIRA et al., 2013) and k-nearest neighbors (MANSUY et al., 2014) have been widely used in precision agriculture.

The objective of this research was to determine the soil moisture from images obtained by an unmanned aerial vehicle (UAV) via *machine learning algorithms*, such as support vector regression (SVR), random forests (RF) and artificial neural networks (ANNs), with the water tension in the soil taken as a reference.

4 MATERIALS AND METHODS

4.1 Study area

The study was conducted on a 174 m² plot located in the experimental area of the Department of Rural Engineering and Socioeconomics of the Faculty of Agricultural Sciences (FCA) of São Paulo State University "Júlio de Mesquita Filho" (Unesp) in Botucatu, SP, as shown in Figure 1.

Figure 1. Location of the study area.
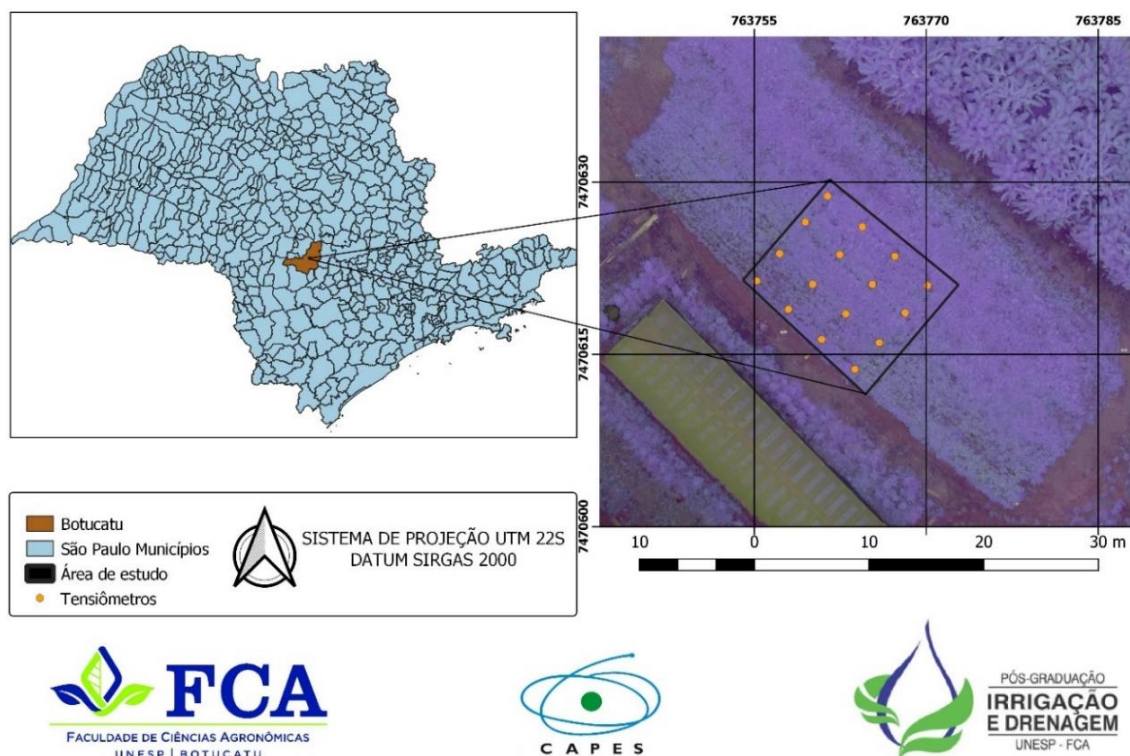
The soil in the area is characterized as a clayey red nitrate, according to the Pedological Map of the State of São Paulo (ROSSI, 2017). According to the Köppen and Geiger (1928) classification, the climate of the region is humid subtropical, Cfa, with two well-defined seasons: hot, humid summers and dry winters. The relief is flat with a 1% slope. The plot was planted with a bean (*Phaseolus vulgaris* L.), variety Dama, with a 91-day cycle. Irrigation was provided via conventional sprinkler irrigation with a 12 × 12 m sprinkler overlap.

Figure 2 shows the equipment used for the aerial survey. The UAV used was a Phantom 3® Professional (manufacturer: SZ DJI Technology Co., Shenzhen, Guangdong, China). The flight was conducted autonomously at an altitude of 120 m, with 60% lateral overlap and 70% frontal overlap.

The sensor used was a MAPIR Survey 3 W (manufacturer: MAPIR, Peau Products, Inc., CA, USA). The MAPIR camera is a modified camera with a filter for recording in the near-infrared region (850 nm) and recording in the red (660 nm) and green (550 nm) regions. The camera has a resolution of 12 megapixels (4032 × 3024) and produces images in JPG (8-bit) and RAW (12-bit) formats. The standard aperture, white balance, ISO sensitivity (International Standards Organization), shutter aperture, and exposure settings were used, as recommended by the manufacturer. Currently, this sensor is the cheapest on the market ($\cong$ R$ 6,000) compared with other multispectral cameras available, such as the Parrot Sequoia ($\cong$ R$ 30,000) and Micasense RedEdge-M ($\cong$ R$ 50,000).

Figure 2. UAV and calibration panel.

A radiometric calibration panel was used to convert the digital numbers into reflectances. Preflight images were collected for calibration, and the captured images were processed via Mapir camera control *software* v. 10/16/2019. For geometric calibration, a Kronos 200 RTK GNSS receiver (manufactured by Horizon, Survey Instruments Ltd., Singapore) was used, and five control points were collected throughout the area. To estimate soil moisture, two surveys were conducted: the first on January 18th and the second on January 20th, both of which were conducted in January 2020.

The instruments used for monitoring soil water tension are shown in Figure 3. A battery of 16 tensiometers installed at a depth of 20 cm was used. The water in the tensiometers was maintained at a constant level after the measurements. Tension readings were taken via a digital needle tensiometer (manufactured by Hidrodinâmica Tensiômetros, Piracicaba, Brazil). The soil samples were collected to obtain and calculate the retention curve via pressure chamber methodology. The tension values were subsequently adjusted via the van Genuchten (1980) model to obtain the values of the soil volumetric water content (cm^3/cm^3). The data were then multiplied by 100 to obtain the soil moisture content as a percentage.

Figure 3. Tensiometer and digital needle tensiometer.

4.2 Support Vector Regression

Support vector machine algorithms can also be applied to regression problems by introducing an alternative loss function (SMOLA, 1996). *Support vector regression* (SVR) algorithms are a generalization of the classification problem found in *support vector machine classifiers*. In these algorithms, errors are fixed so that points within the confidence interval are discarded for regression. The optimization criterion penalizes data points in which the y values differ from f(x) by more than the error ε.

The generic SVR estimation function can be described as Equation (1):

$$f(x) = \big(\omega * \Phi(x)\big) + b \qquad (1)$$

where Φ is a nonlinear transformation to a higher-dimensional space. The objective is to find a value of ω and b such that the values of *x* can be determined while minimizing the regression risk (R_{reg}) represented by Equation 2 (WU; HO; LEE, 2004).

$$R_{reg} = C \sum_{i=0}^{l} \Gamma \left(f(x_i) - y_i \right) + \frac{1}{2} \|w\|^2 \qquad (2)$$

where Γ () is a cost function, C is a constant, and the vector *w* can be written in terms of the data points (Equation 3):

$$w = \sum_{i=1}^{l} (a_i - a_i^*) \Phi(x_i) \qquad (3)$$

The transformation product Φ can be estimated via the function *k(x ᵢ, x), which is* called the kernel function. The radial basis function (RBF) is the most widely used kernel and can be defined according to Equation (4):

$$RBF = exp\{-\gamma |x - x_i|\} \qquad (4)$$

The most commonly used cost function in the literature is the e-insensitive function. It determines a tube bounded by the support vectors for cutting. It is solved as

a function of an acceptable error value on the basis of the data (Equation (5)).

$$\Gamma(f(x) - y) = |f(x) - y| \geq \varepsilon \qquad (5)$$

More details on SVR theory can be found in Smola and Scholkpof (2004) and Vapnik (1998).

4.3 Random Forests

The random forest algorithm is a supervised learning algorithm that uses ensemble learning for classification and regression. In the algorithm, several decision trees are randomly created, forming something similar to a forest, where each tree is used to define the final result. There is a collection of predictor trees $h(x; \theta_k)$, $k = 1..$, K, where x represents the observed input vector of length p with a random vector **X** and θ_k are identically distributed random vectors (SEGAL, 2004).

The mean squared error of generalization for any numerical predictor h (**x**) is described according to Equation (6) (BREIMAN, 2001):

$$E_{X,Y}(y - h(\boldsymbol{X}))^2 \qquad (6)$$

The random forest predictor is formed by taking the average over k of the trees $\{h(\mathbf{x}, \theta_k)\}$ as the number of trees tends to infinity (Equation 7):

$$E_{X,Y}(y - méd_k h(\boldsymbol{X}, \theta_k))^2 \rightarrow E_{X,Y}(y - E_\theta h(\boldsymbol{X}; \theta))^2 \qquad (7)$$

The quantity on the right is the prediction (generalization) error for the *random forest* designated PE^{f*}. The prediction error of an individual tree $h(X;\theta)$ can be defined as (Equation 8):

$$PE_t^* = E_\theta E_{X,Y}(y - h(\boldsymbol{X}; \theta))^2 \qquad (8)$$

Assuming that for all θ, the tree has no bias ($EY = E_x h(\boldsymbol{X}, \theta)$). Then, we have Equation 9 (SEGAL, 2004):

$$PE_f^* \leq \bar{p} PE_t^* \qquad (9)$$

where $\bar{p}$ is the weighted correlation between the residuals y – $h(\boldsymbol{X};\theta)$ and y- $h(\boldsymbol{X};\theta')$ for θ, θ', which are randomly distributed independent vectors .

Random *forests* work by constructing multiple decision trees during the training process and outputting *the* average value predicted by the individual trees. Each tree extracts a random sample from the set, adding an additional element of randomness and preventing *overfitting*. The trees are run in parallel, with no interaction between them when new trees are constructed. This is classified as a *bagging algorithm*, not a boosting algorithm such as neural networks.

4.4 Artificial Neural Network

Artificial neural networks, or artificial neural networks, are algorithms that use an approach similar to the structure of the human brain for decision making (MCCULLOCH; PITTS, 1943; ROSENBLATT, 1962; BISHOP, 1995). The independent input variables x_i (i = 1, . . ., d) are transformed into a dependent output set $y_1,...., y_i$. The first stage of the transformation is performed by the neuron or *perceptron*. The *inputs* are multiplied by a weight parameter w_i that simulates the synaptic weights in biological neural networks. All weighted inputs are summed to obtain a total input (Equation 10):

$$a = \sum_{i=1}^{d} w_i x_i + b \qquad (10)$$

A bias b is added to provide a mechanism for including other influences,

which is typically set to 1. The learning process begins with some arbitrary weight vector, without loss of generalization; we can assume ~~this to be the zero~~ vector (Bishop, 1995). An activation function will define the acceptable threshold for passing the value of the weighted sum plus the bias value.

-linear activation function f () that will define, on the basis of the threshold, whether it will be activated or not (Equation 11).

$$z = f(a) \qquad (11)$$

Networks typically use the sigmoid function to calculate thresholds. Because this type of network operates through neurons, from *inputs* to *outputs,* it is called *feed-forward* (Bishop, 1995). The simplest network structures have only one neuron, as shown in Figure 4. However, the most commonly used structure is called *the multilayer perceptron* (MLP), as it combines several layers of neurons.

Figure 4. Mathematical model of an artificial neuron. Source: Adapted from Haykin (2009).
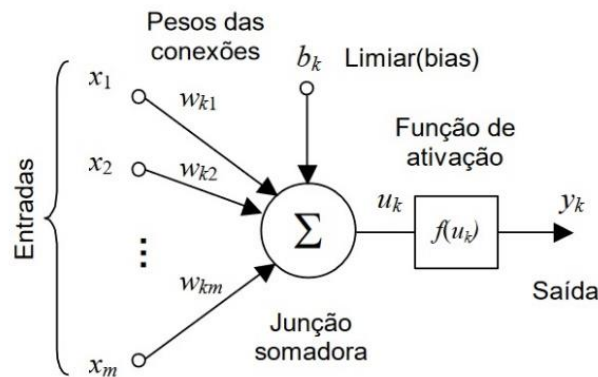
The multilayer perceptron structure is characterized by an input layer, intermediate layers called hidden layers, and an output layer. As the learning process progresses through the MLP, the neurons in the hidden layer gradually begin to discover the salient features that characterize the training data (HAYKIN, 2009).

The network learns via the *backpropagation algorithm*, which uses a technique called gradient descent. This algorithm adjusts the weights on the basis of their derivative with respect to the error. The network is initialized, and errors are calculated at the end of the process; they are then propagated back to the initial layers for

weight adjustment. The process repeats several times until the combination of synaptic weights that results in the smallest error is found, at which point the model converges.

This technique measures the error and the rate of change of the error. This leads to large changes in the larger error, and as the slope decreases as it approaches a minimum, the changes in the weights decrease (PUJOL; PINTO, 2011). The mean of all squared errors (E) for the output is computed to aid in the derivative. The descent is based on a gradient in the error for the entire dataset according to Equation (12) (GROSSAN; ABRAHAM, 2011):

$$\Delta w_{ij}(n) = -\eta * \frac{\delta E}{\varepsilon_{wij}} + a * \Delta w_{ij}(n-1) \qquad (12)$$

where η * is the learning rate and a * is the momentum.

The use of this algorithm has already been reported in numerous studies whose main objective was to obtain an estimate of soil moisture (JIANG; COTTON, 2004; AHMAD; KALRA; STEPHEN, 2010; KORNELSEN; COULIBALY, 2014; HASSAN-ESFAHANI et al., 2015).

4.5 Modeling strategy

buffers with a radius of 0.5 m around the tensiometer locations were used to extract the mean reflectance values for each band via the zonal statistics tool in *QGIS 3.4. These data were used as inputs for machine learning* algorithms. The independent variables were the reflectances in the green, red, and near-infrared bands and the normalized difference vegetation index (NDVI). ~~The~~ dependent variable ~~was~~ the soil moisture recorded by each tensiometer in the two surveys (January 18 and 20, 2020). The input matrix had a 32 × 4 format.

The data were split into 70% for training and 30% for testing predictions. The data were subsequently normalized via the *standard scaler* (Equation (13)), which standardizes the samples by removing the mean and leaving the unit variance with a Gaussian normal distribution. This process is a requirement for *machine learning algorithms* to ensure that the data are on the same scale.

$$z = \frac{(x-u)}{s} \qquad (13)$$

where *x* is a sample, *u* is the mean of the training samples and *s* is the standard deviation.

The algorithms used in this study were implemented via Python 3.6 through the *Jupyter Notebook user interface*, along with the Pandas, Numpy, and Scikit-Learn libraries. Regarding the hyperparameters selected for the estimators, the *kernel radial basis function was used in the SVR algorithm.*

For random forests (RFs), 100 decision trees were used, and the optimization criterion was the *'mean square error'*. For artificial neural networks (ANNs), an MLP architecture was used, initialized with two hidden layers with 100 neurons each with the ReLU activation function. The cost function used was the "*square error*", and the learning rate was set to 0.001. The selected optimizer was 'adam', with the output layer without an activation function, thus generating a continuous number.

4.6 Statistical metrics

To evaluate the performance of the *machine learning algorithms* used to determine the soil moisture estimate, the following statistical metrics were used: *root mean square error* (RMSE) (Equation 14), *mean absolute error (MAE)* (Equation 15), *mean absolute percentage error* (MAPE) (Equation 16) and coefficient of determination (R^2) (Equation 17).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \qquad (14)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \qquad (15)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - x_i}{x_i}\right| * 100 \qquad (16)$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (17)$$

where y_i represents the observed soil moisture values, x_i represents the simulated values, *n* represents the number of observations and $\bar{y}$ represents the mean of the observations.

5 RESULTS AND DISCUSSION

5.1 Soil moisture and reflectance

Figure 5 shows the correlations between the spectral bands (green, red, and near-infrared), NDVI, and soil moisture indirectly estimated by tensiometers. A negative correlation is observed between the red band and soil moisture. When the plant is not stressed, radiation in the red band is reflected with lower intensity because of the absorption of chlorophyll by carotenes and xanthophylls. Therefore, areas with high soil moisture contents have lower reflectances in this range of the electromagnetic spectrum.

Figure 5. Scatter plots demonstrating the correlation between soil moisture and spectral bands.
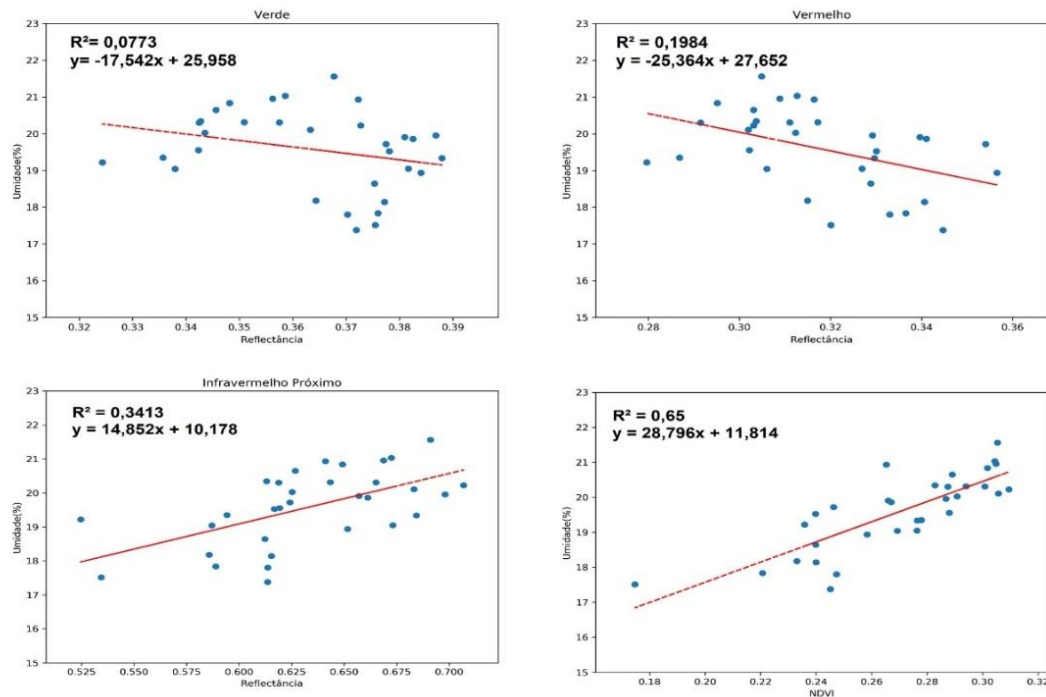
A positive correlation between the near-infrared band and soil moisture can also be identified. This occurs because where there is greater water availability, the plant has greater vegetative development, increasing near-infrared reflectance due to the higher leaf area index (LAI). The green band did not show a significant correlation with soil moisture at the analyzed depth. These results are similar to those reported by Aboutalebi et al. (2019), who reported the same trend of correlation between soil moisture and this spectral band in soil layers from 45 cm deep.

Notably, there was a strong positive correlation between the NDVI and soil moisture (Figure 5). The NDVI is related to the amount of moisture in the soil, as a plant without water restrictions can reach full vegetative vigor, presenting high NDVI values, whereas areas with low NDVI values, generally associated with a greater presence of exposed soil in the pixels, may indicate that the plants are experiencing water stress due to low soil water availability (lower moisture). The MAPIR camera tends to present low NDVI values, as reported by Gomes et al. (2021).

5.2 Performance of machine learning estimators

Figure 6 presents the humidity estimated by the *machine learning algorithms* and the humidity observed in the tensiometers for the training and test sets. All

algorithms performed satisfactorily, with an RMSE < 1 vol. % in estimating soil moisture.

Figure 6. Scatter plots of soil water tension values observed by tensiometers versus values estimated by *machine learning algorithms*.
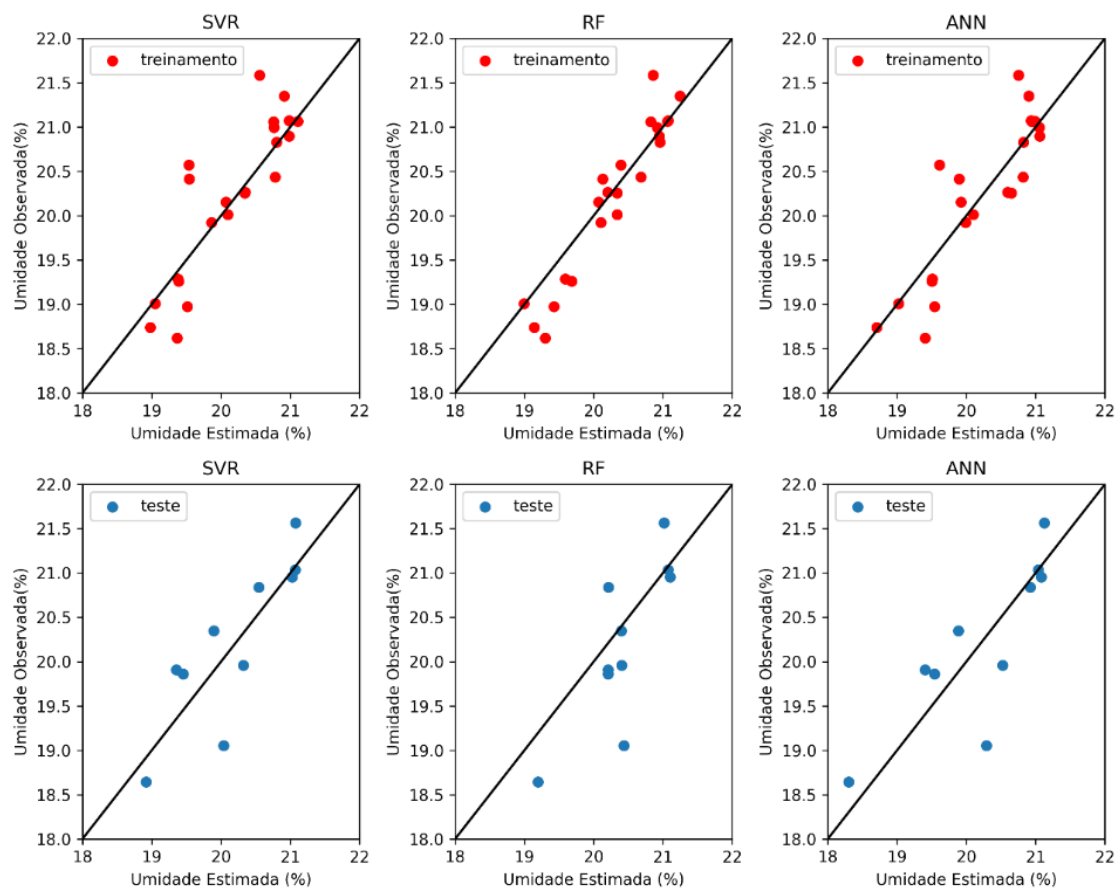

Table 2 presents a summary of the statistics evaluated via the test data. The SVR algorithm performed better than the RF and ANN algorithms did. This can be confirmed by the distribution of points in relation to the 1:1 line, where clustering was greater in the SVR algorithm. The SVR presented an RMSE of 0.46 vol. %, and the MAPE value was 4.59%. The regression fit was 0.71, and the MAE was 0.39 vol. %. The performance of the ANN algorithm was similar to that of SVR, with an RMSE of 0.54 vol. % and a MAPE of 4.23%; the R^2 was 0.60, and the MAE was 0.40 vol. %. The RF presented an RMSE of 0.57 vol. % and a MAPE of 4.23%. The R^2 was 0.55, with an MAE of 0.42 vol. %.

Table 1. Summary statistics of observed versus estimated values of percentage soil moisture.

	RMSE[1] (vol %)	MOTHER (vol %)	MAPE (%)	R²
SVR[2]	0.46	0.39	4.59	0.71
RF	0.57	0.42	4.23	0.55
ANN	0.54	0.40	4.75	0.60

Overall, the models accurately estimated soil moisture via the green, red, near-infrared, and normalized difference vegetation index (NDVI) wavelengths. Very low error values are typically associated with overfitting, but when the algorithms are evaluated on validation (test) data, the low R2 values demonstrate that the fit is not perfect, a key characteristic of *overfitting.* where R^2 is equal to or very close to 1. However, the humidity values did not present a significant difference across the surveys (1.3%), which may explain the low RMSE values.

The results presented are similar to those reported by Araya et al. (2021), who obtained MAE values of 3.77 vol.% when the same algorithms employed in this work were applied to multispectral data collected by the Parrot Sequoia sensor (manufactured by Parrot SA, Paris, France) in six surveys with 406 soil moisture samples. Ge et al. (2019) also reported similar results (RMSE = 1.47 vol. %) when the RF and *extreme learning machine* algorithms were used on 70 soil moisture samples collected in a survey with a hyperspectral camera. This demonstrates the potential of using *machine learning methods* on data collected via remote sensing and their application in monitoring soil moisture content, which, consequently, allows for adequate irrigation management, presenting itself as a viable option for precision agriculture.

5.3 Soil moisture maps

Figure 7 shows the maps generated by the *machine learning algorithms* in the surveys conducted on January 18th and 20th. All the maps show lower moisture values in the lower left corner, which is consistent with the data recorded by the tensiometers. The RF algorithm smooths the data, generating a more uniform map. The SVR algorithm, on the other hand, generated a soil moisture map with large spatial variation for both surveys. The ANN algorithm performed worst in estimating soil moisture, presenting greater generalization of values, and even allowing for the identification of *clusters* with more defined positions.

---

[1]Root Mean Square Error (RMSE); Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE), Coefficient of determination (R^2).
[2]Support Vector Regression (SVR); Random Forests (RF) and Artificial Neural Network (ANN).

Figure 7. Soil moisture maps generated by *machine learning algorithms* in both surveys.
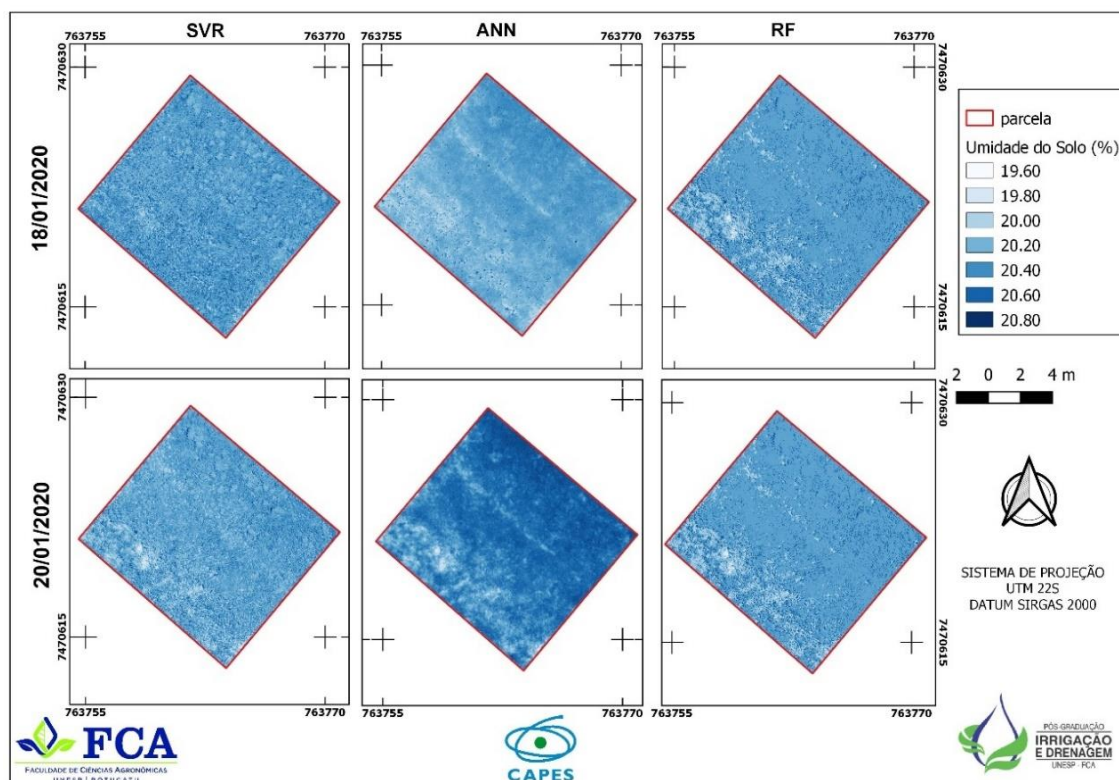
The maps were generated at high resolution with a 12 cm pixel size, which may not be technically feasible for irrigation management. However, the data can be resampled via the *downsampling technique*, which allows for reduced spatial resolution. Another alternative is to conduct flights at altitudes above 120 m, which is more suitable for surveys in large areas such as those occupied by central pivot irrigation systems. However, the objective of this work was to demonstrate the accuracy of *machine learning algorithms* in estimating soil moisture via data collected by a low-cost optical sensor. The ability to collect data at any time interval via a UAV, combined with the pattern recognition capability of *machine learning* algorithms, constitutes an excellent tool for remote soil moisture data acquisition.

6 CONCLUSIONS

This study demonstrated the usefulness of *machine learning algorithms* for estimating soil moisture from high-resolution multispectral images collected by UAVs. The red and near-infrared bands were the most strongly correlated with the soil water tension at a depth of 20 cm. The performances of the SVR and ANN algorithms were similar, with little difference in the regression fit. The best algorithm for estimating soil moisture, on the basis of the data analyzed, was SVR, which presented an RMSE of 0.46 vol.% in estimating soil moisture in the test data and an R^2 of 0.71. The maps generated by the algorithms demonstrate the high spatial variability of soil moisture and can be used for monitoring it at any time scale. The technique presented here is limited only by climatological conditions that may impede data collection via a UAV.

The performance of *machine learning algorithms*, especially neural networks, is influenced by the size of the training data. *Overfitting may occur* in smaller datasets, which limits the results presented here. However, it is important to note that new data collection campaigns can be conducted, and these data can be used to retrain the models, thus increasing their accuracy.

This study provides evidence that *machine learning* can be used to indirectly estimate soil water content, making it possible to obtain highly accurate results even when low-cost cameras attached to unmanned aerial vehicles are used.

7 ACKNOWLEDGMENTS

8 REFERENCES

ABOUTALEBI, M.; ALLEN, LN; TORRES-RUA, AF; MCKEE, M.; COOPMANS, C. Estimation of soil moisture at different soil levels using machine learning techniques and unmanned aerial vehicle (UAV) *multispectral* imagery **"** .

AHMAD, S.; KALRA, A.; STEPHEN, H. Estimating soil moisture using remote sensing data: A machine learning approach. **Advances in Water Resources** , Iowa City, vol. 33, p. 69-80, 2010.

AITKENHEAD, MJ; COULL, M.; TOWERS, W.; HUDSON, G.; BLACK, HIJ Prediction of soil characteristics and color using data from the National Soils Inventory of Scotland. **Geoderma** , Amsterdam, v. 200/201, p. 99-107, 2013.

ARAYA, SN; FRYJOFF-HUNG, A.; ANDERSON, A.; VIERS, JH; GHEZZEHEI, TA Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. **Hydrologic Earth Systems Science** , Gottingen, v. 25, p. 2739-2758, 2021.

ARRUDA, LEV; FIGUEIRÊDO, VB; LEVIEN, SLA; MEDEIROS, JF Development of a digital tensiometer with data acquisition and storage system. **Irriga** , Botucatu, v. 1, n. 1, p. 11-20, 2017.

BISHOP, CM **Neural Networks for Pattern Recognition** . Oxford: Oxford University Press, 1995.

BREIMAN, L. Random forests. **Machine Learning** , New Jersey, v. 45 , p. 5-32, 2001.

BRITO, AS; LIBARDI, PL; MOTA, JCA; MORAES, SO Tensiometer performance with different reading systems. **Brazilian Journal of Soil Science** , Viçosa, v. 33, n. 1, p. 17-24, 2009.

BRITO, AS; LIBARDI, PL; MOTA, JCA; KLEIN, VA Diurnal-nocturnal variation of matric potential and total soil water potential gradient . **Brazilian Journal of Soil Science** , Viçosa, v. 38, n. 1, p. 128-134, 2014.

CAMARGO, AP; GROHMANN, F.; CAMARGO, MBP Simple direct-reading tensiometer. **Brazilian Agricultural Research** , Brasília, DF, v. 17, n. 12, p. 1763-72, 1982.

DOBRIYAL, P.; QURESHI, A.; BADOLA, R.; HUSSAIN, SA A review of the methods available for estimating soil moisture and its implications for water resource management. **Journal of Hydrology** , Amsterdam, v. 458-459, p. 110-117, 2012.

DUKES, MD; ZOTARELLI, L.; MORGAN, KT Use of irrigation technologies for vegetable crops in Florida. **HortTechnology** , Alexandria, v. 20, no. 1, p. 133-142, 2010.
FREITAS, WA; CARVALHO, JA; BRAGA, RA; ANDRADE, MJB Irrigation management using alternative soil moisture sensor. **Brazilian Journal of Agricultural and Environmental Engineering** , Campina Grande, v. 16, n. 3, p. 268-274, 2012.

GE, X.; WANG, J.; DING, J.; CAO, X.; ZHANG, Z.; LIU, J.; LI, X. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. **Environmental Science** , Amsterdam, v. 7, n. 6929, p. 1-27, 2019.

GOMES, APA; QUEIROZ, DM; VALENTE, DSM; PINTO, FAC; ROSAS, JTF Comparing a single-sensor camera with a multisensor camera for monitoring coffee crops using unmanned aerial vehicles. **Agricultural Engineering** , Jaboticabal, v. 41, n. 1, p. 87-97, 2022.

GROSSAN, C.; ABRAHAM, A. **Intelligent Systems** : A modern approach. Berlin: Springer, 2011.

HASSAN-ESFAHANI, L.; TORRES-RUA, A.; JENSEN, A.; MCKEE, M. Assessment of surface soil moisture using high-resolution multispectral imagery and artificial neural networks. **Remote Sensing** , Basel, v. 7 , no. 3, p. 2627-2646, 2015.

HAYKIN, S. **Neural Networks** : A Comprehensive Foundation. 2nd ed. New Jersey: Prentice Hall, 1999.

HUISMAN, JA; HUBBARD, SS; REDMAN, J.D.; ANNAN, AP Measuring soil water content with ground penetrating radar: a review. **Vadose Zone Journal** , Madison, vol. 2, no. 4, p. 476-491, 2003.

JENSEN, JR **Remote sensing of the environment** : an earth resource perspective. 2nd ed. New Delhi: Pearson Education, 2009.

JIANG, H.; COTTON, W. Soil moisture estimation using an artificial neural network: A feasibility study. **Canadian Journal of Remote Sensing** . Québec, vol. 30, no. 5, p. 827-839, 2004.

KÖPPEN, W.; GEIGER, R. **Klimate der Erde** . Gota: Justus Perthes, 1928.

KOVAČEVIC, M.; BAJAT, B.; GAJIĆ, B. Soil type classification and estimation of soil properties using support vector machines. **Geoderma** , Amsterdam v. 154, n. 3/4, p. 340-347, 2010.

KORNELSEN, K.C.; COULIBALY, P. Root-zone soil moisture estimation using data-driven methods. **Water Resources Research** , New Jersey, vol. 50, p. 2946-2962, 2014.

LIBARDI, PL **Soil water dynamics** . São Paulo: EDUSP, 2005.

MANSUY, N.; THIFFAULT, E.; PARÉ, D.; BERNIER, P.; GUINDON, L.; VILLEMAIRE, P.; POIRIER, V.; BEAUDOIN, A. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. **Geoderma** , Amsterdam, v. 235/236, p. 59-73, 2014.

MCCULLOCH WS; PITTS. W. A logical calculus of the ideas immanent in nervous activity, **Bulletin of Mathematical Biophysics** , New York , v. 5, p. 115-133, 1943.

MONTESANO, FF; SERIO, F.; MININNI, C.; SIGNORE, A.; PARENTE, A.; SANTAMARIA, P. Tensiometer-based irrigation management of subirrigated soilless tomato: effects of substrate matrix potential control on crop performance. **Frontiers in Plant Science** , Lausanne, vol. 6, no. 1150, p. 1-11, 2015.

NEW, EML **Remote Sensing** : principles and applications. Sao Paulo: Edgard Blucher, 2008.

OR, D. Who invented the tensiometer? **Soil Science Society of America Journal** , New Jersey, vol. 65, no. 1, p. 1-3, 2001.

PONZONI, FJ; SHIMABUKURO, YE **Remote Sensing in the study of vegetation** . São José dos Campos: Parentheses, 2010.

PRIORI, S.; BIANCONI, N.; CONSTANTINI, EAC Can γ -radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine learning methods. **Geoderma** , Amsterdam, v. 226/227, p. 354-364, 2014.

PUJOL, JCF; PINTO, JMA A Neural-network approach to fatigue life prediction. **International Journal of Fatigue** , Amsterdam, v. 33, n. 3, p. 313-322, 2011.

ROSA, R. **Introduction to remote sensing** . 7th ed. Uberlândia: EDUFU, 2009.

ROSENBLATT, F. **Principles of neurodynamics** : perceptrons and the theory of brain mechanisms. New York: Spartan Books, 1962.

ROSSI, M. **Pedological map of the state of São Paulo** : revised and expanded . São Paulo: Instituto Florestal, 2017.

SEGAL, MR **Machine Learning Benchmarks and Random Forest Regression** . San Francisco: University of California, 2004. Available at: h ttps://escholarship.org/uc/item/35x3v9t4. Accessed on: February 9 , 2022.

SILVEIRA, CT; OKA-FIORI, C.; SANTOS, LJC; SIRTOLI, AE; SILVA, CR; BOTELHO, M.F. Soil prediction using artificial neural networks and topographic attributes. **Geoderma** , Amsterdam, v. 195/196, p. 165-172, 2013.

SMOLA, J. **Regression estimation with support vector learning machines** . Dissertation (Master in Physics) – Technische Universitat at Munchen, Munich, 1996.
SMOLA, AJ; SCHOLKOPF, B. **A tutorial on support vector regression** . London: Royal Holloway College, 2004.
TEIXEIRA, AS; COELHO, SL Development and calibration of an automatic reading electronic tensiometer. **Agricultural Engineering** , Jaboticabal, v. 25, n. 2, p. 367-376, 2005.

THALHEIMER, M. A low cost electronic tensiometer system for continuous monitoring of soil water potential. **Journal of Agricultural Engineering** , Pavia, v. 44, no. 3, p. 114-119, 2013.

VAN GENUCHTEN, MT A closed form equation for predicting the hydraulic conductivity of unsaturated soils. **Soil Science Society of America Journal** , Madison, vol. 44, n. 5, p. 892-898, 1980.

VAPNIK, V. **Statistical Learning Theory** . New York: Springer, 1998.

WU, CH; dHO, JM; LEE, DT Travel-time prediction with support vector regression. **IEEE Transactions on intelligent transportation systems** , Blacksburg , v. 5, no. 4, p. 276-281, 2004.