Machine learning algorithms applied to the forecasting of crop water stress indicators

Manuel A. Andrade

Oak Ridge Institute for Science and Education, 100 ORAU Way, Oak Ridge, TN 37830, USA, Alejandro.Andrade@ars.usda.gov

Steven R. Evett

USDA-Agricultural Research Service, P.O. Drawer 10, 2300 Experimental Station Road, Bushland, TX, 79102, USA, Steve.Evett@ars.usda.gov

Susan A. O'Shaughnessy

USDA-Agricultural Research Service, P.O. Drawer 10, 2300 Experimental Station Road, Bushland, TX 79102, USA, Susan.O'Shaughnessy@ars.usda.gov

Abstract. Recent advances made by scientists with the USDA-Agricultural Research Service (ARS) can provide farmers with irrigation scheduling tools based on crop stress indicators to assist the management of Variable Rate Irrigation (VRI) center pivot systems. These tools were integrated into an Irrigation Scheduling Supervisory Control and Data Acquisition System (ISCCADAS). The ISSCADAS automates the collection of data from a network of wireless infrared thermometers (IRTs) distributed on a center pivot's lateral and in the field irrigated by the center pivot, as well as data from a wireless soil water sensor network in the field and a microclimate weather station located at the pivot point. This study analyzes the use of Artificial Neural Networks, one of the most popular machine learning algorithms, for the forecasting of canopy temperatures obtained by a wireless network of IRTs mounted on a three-span VRI center pivot irrigating corn near Bushland, TX, during the summer of 2017. Two case studies were conducted for this purpose using data collected from periodic scans of the field performed during the growing season by running the pivot dry. In the first case, data from the first three scans were used to train an Artificial Neural Network (ANN). Canopy temperatures estimated using the ANN were then compared against canopy temperatures measured by the network of IRTs during the fourth scan. In the second case, data from the first six scans were used to train ANNs. Then ANN predicted canopy temperatures were compared against canopy temperatures measured during the seventh scan. The Root of the Mean Squared Error (RMSE) of ANN predictions in the first case ranged from 1.04 °C to 2.49 °C, whereas the RMSE of ANN predictions in the second case ranged from 2.14 °C to 2.77 °C. To assess the impact of ANN accuracy on irrigation management, estimated canopy temperatures were fed to a plantstress based irrigation scheduling method and the resulting prescription maps were compared against prescription maps obtained by the same method using the canopy temperatures measured by the network

of IRTs. In the first case no difference was found between both prescription maps. In the second case only one plot (out of 26) was assigned a different prescription. Results of this study suggest that machine learning techniques can be used to assist the ISSCADAS in situations where canopy temperatures cannot be measured by the network of IRTs due to poor visibility conditions, or because the center pivot cannot traverse the field within a reasonable amount of time.

Keywords: machine learning, metamodeling, center pivot irrigation, variable rate irrigation, irrigation scheduling, sensors.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

The mention of trade names of commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Introduction

Agriculture, like many other economic sectors, is facing a rapid transformation resulting from the use of data analytics (Pham and Stack, 2018). Machine learning is a data analysis technique that can be used to solve a wide range of problems, since it uses previous data to train computational 'machines' that can make predictions on new data with no human intervention. One of the most popular types of machine learning algorithms is the Artificial Neural Network (ANN), which is capable of handling complex associations between inputs and outputs. This characteristic makes the ANN particularly well fitted for solving agricultural problems where no known deterministic model may be available to simulate the complex interactions occurring between the many components involved in an agricultural system. In agricultural engineering, ANNs have been used to, among other applications, predict soil properties (Schaap et al., 1998; Farifteh et al., 2007), estimate crop yield (Kaul et al., 2005), model greenhouse environmental variables (Ferreira et al., 2002; He and Ma, 2010), and analyze images for the discrimination of crop and weeds (Yang et al., 2000; Kavdir, 2004).

This study uses ANNs to forecast canopy temperatures obtained by a wireless network of infrared thermometers (IRTs) mounted on a three-span VRI-center pivot irrigating corn near Bushland, TX, during the summer of 2017. The network of IRTs is part of an Irrigation Scheduling Supervisory Control and Data Acquisition System (ISSCADAS) patented by scientists with the USDA-Agricultural Research Service (ARS) (Evett et al., 2014). The ISSCADAS collects data from soil water, plant, and weather sensing systems and feeds those data to computerized irrigation scheduling algorithms based on plant stress to generate site-specific prescription maps. The ISSCADAS also integrates hardware functions to manage the submission of prescription maps for their application by VRI center pivot systems operated with a Pro2 control panel (Valmont Industries Inc., Valley NE). A software package, named ARS-Pivot (ARSP), was developed to simplify the operation of the ISSCADAS (Andrade et al., 2015, 2017). The main objective of this study is to analyze the feasibility of using ANNs to predict future canopy temperatures based on those previously measured by the aforementioned network of IRTs as the center pivot moves through the field. Results obtained from this study can provide guidelines for the future integration of canopy temperature forecasting using ANNs as one of the tools available in ARSP. The availability of such a tool can add redundancy to the ISSCADAS so that site-specific prescription maps can be generated even if a direct measurement of canopy temperatures is not possible due to poor visibility conditions, or because the center pivot cannot traverse the entire field within a reasonable amount of time.

Methodology

In the summer of 2017, the ISSCADAS and the ARSP software were used for the integrated irrigation management of a three-span center pivot (131 m) located at the USDA-ARS Conservation and Production Research Laboratory, near Bushland, TX. The center pivot was equipped with a Pro2 control panel and a commercial VRI system (Valmont Industries Inc., Valley NE). A midseason corn hybrid, Dupont Pioneer P1151AM, was planted on May 15, day of year (DOY) 135. Experimental plots used in this study were located within the six outermost sprinkler zones in the field (Fig. 1).

VRI zone control was used for the North-Northwest (NNW) side of the field, which was divided into six control sectors of 28° each and six concentric control zones with a width of 9.14 m (30 ft) each, for a total of 36 management zones, each of which was considered an experimental plot. Plots were organized using a Latin square design (Fig. 1). VRI speed control was used for the South-Southeast (SSE) side of the field, which was divided into eight control sectors of 20° each and a single concentric control zone with a width of 54.9 m, for a total of 8 management zones, each of which was considered an experimental plot

(Fig. 1). The irrigation of plots in the NNW side was triggered by either the integrated Crop Water Stress Index (iCWSI) method described by O'Shaughnessy et al. (2017) or by weekly neutron probe (NP) (model 503DR1.5, Instrotek, Campbell Pacific Nuclear, Concord, CA) measurements. Each of these plots was assigned one of the following irrigation levels: 80%, 50%, or 30% of full irrigation. Full irrigation was defined as the irrigation required to return soil water content in the root zone to field capacity. The combination of irrigation scheduling methods (2) and irrigation levels (3) resulted in six treatments with six replicates per treatment (Fig. 1). Plots irrigated with the iCWSI method are labeled in Fig. 1 as C80, C50, or C30, where 'C' stands for iCWSI-based control and numbers correspond to irrigation levels. Similarly, plots irrigated with the NP method are labeled in Fig. 1 as U80, U50, or U30, where 'U' indicates that irrigation scheduling is controlled by the user.

Plots in the SSE side were all assigned a single irrigation level of 80%; their irrigation was triggered by either the iCWSI method, or by a hybrid method using the iCWSI method and an average soil water depletion in the root zone (SWDr) calculated using sets of three time domain reflectometer (TDR) sensors (model 315, Acclima, Meridian, ID) buried at depths of 15 cm, 30 cm, and 45 cm. The hybrid method used a two-step approach for irrigation scheduling. During the first step, the SWDr was compared against pre-determined lower and upper SWDr thresholds. No irrigation was assigned if the SWDr was lower than 0.1 (lower threshold) and an irrigation depth of 30.5 mm (1.2 in) was assigned if the SWDr was higher than 0.5 (upper threshold). If the SWDr fell between these values, the iCWSI method was used during a second step to determine its prescription. Plots irrigated with the hybrid method are labeled in Fig. 1 as H80.

The iCWSI method is based on calculation of the theoretical Crop Water Stress Index (CWSI) (Jackson et al., 1981) at discrete intervals during daylight hours. CWSI values were calculated for each location *x* in the field at time interval *t* using the normalized difference between the crop canopy temperature in the location and the air temperature at time *t*. Additional details of the iCWSI method and the formulas used for its calculation can be found in O'Shaughnessy et al. (2013) and O'Shaughnessy et al. (2017). Air temperature and other relevant weather parameters (relative humidity, solar irradiance, wind speed, and wind direction) were sampled every 5 s and averaged and stored every minute at a weather station (Campbell Scientific, Logan, UT) located next to the pivot point. Crop canopy temperatures were measured at two fixed locations in the field using wireless infrared thermometers (IRTs) (model SapIP-IRT, Dynamax Inc., Houston, TX) to provide a reference canopy temperature for a well-watered crop (Fig. 1). A network of 12 wireless infrared thermometers (IRTs) (model SapIP-IRT, Dynamax Inc., Houston, TX) was mounted on the center pivot to measure canopy temperatures inside the experimental area (Fig. 1). The IRTs were located forward of the drop hoses, at an oblique angle from nadir. The

average of data collected from two IRTs with opposing views of a sprinkler control zone was the primary datum every minute for each sprinkler zone. Pairs of IRTs arranged in such a way are referred hereinafter as IRT groups.

Scans of the field were performed periodically through the growing season by running the center pivot dry. Weather data and canopy temperatures—measured by the network of stationary IRTs in the field and on the center pivot—collected during scans were used to train ANNs to estimate average canopy temperatures obtained by a given IRT group, i.e., by a pair of IRTs with opposing views of a sprinkler zone. Two case studies were conducted to analyze the feasibility of using ANNs for this purpose. In the first case, six types of ANNs (one for each of the six IRT groups located on the center pivot) were trained using data collected during the first three scans that took place on June 26 (DOY 177), July 7 (DOY 188), and July 11 (DOY 192). Since the training of ANNs is a semi-random process that yields different results every time, 50 ANNs were trained for each ANN type and the best performing ANN among them was then selected to be used for the forecasting of average canopy temperatures that would be measured by the corresponding IRT group during the following scan (July 12, DOY 193). The accuracy of the best ANN selected for ANN type n was then assessed by predicting average canopy temperatures that would be measured by IRT group n on this date. In the second case, six types of ANNs were trained using data collected during the first six scans that, in addition to the previous dates, took place on July 17 (DOY 198), and July 20 (DOY 201). 50 ANNs were also trained for each ANN type and the best performing ANN was selected to be used for the forecasting of average canopy temperatures that would be measured by the corresponding IRT group during the following scan (July 24, DOY 205).

The typical structure of an ANN, also known as architecture, is composed of at least three layers of nodes (usually referred to as neurons) and the links between these layers (Fig. 2). The first layer is the input layer, the last one is the output layer, and all others are hidden layers. Nodes in these layers are referred to as input neurons, output neurons, and hidden neurons, respectively. ANNs used in this study were three-layered feed-forward networks consisting of sigmoid hidden neurons and linear output neurons. The number of input neurons was 10, corresponding to the number of variables that were considered relevant for the estimation of average crop canopy temperatures estimated by a given IRT group *n* mounted on the center pivot (Fig. 1). These variables were: (1) air temperature measured at time *t* during a scan, (2) relative humidity at time *t*, (3) solar irradiance at time *t*, (4) wind direction at time *t*, (5) wind speed at time *t*, (6) average canopy temperature measured by stationary IRTs at time *t*, (7) irrigation level (%) assigned to the experimental plot *p* being scanned by IRT group *n* at time *t*, (8) irrigation scheduling method assigned to plot *p*, (9) number of days passed since planting at the time of the scan, and (10) cumulative irrigation (including precipitation) received by experimental plot *p*.





ANNs with a single output neuron are expected to be better estimators than ANNs with multiple output neurons (Andrade et al., 2016) and thus a single output neuron was selected for ANNs used in this study. The only output neuron allocated the average canopy temperature measured by a given IRT group mounted on the center pivot (Fig. 2). Using a single output neuron for ANNs in this study offers the additional advantage of allowing ANNs to account for conditions that may be exclusive to a single IRT group, such as scanning a sprinkler zone with a clogged nozzle. The number of neurons in the hidden layer was selected as 12 after running a series of preliminary tests with 2, 4, 6, ..., 18 hidden neurons. These numbers were tested following rules of thumb for the selection of the number of hidden neurons (Heaton, 2008). Additional preliminary tests were performed to determine if the addition of a second hidden layer improved the accuracy of ANNs, but their accuracy did not show to be consistently better

than ANNs with a single hidden layer. The training of ANNs was performed using Matlab's neural network toolbox software (Beale et al., 2011). The Resilient Backpropagation method was selected for the training of ANNs, since it performed consistently better than other training algorithms tested during preliminary runs. The other algorithms tested were Levenberg-Marquardt, Bayesian Regularization, and Conjugate Gradient.

Datasets used for the training of ANNs in the first case study can be represented by an input matrix with dimensions M by N, and an output vector with M elements, where M is the total number of one-minute intervals occurring during the first three scans performed in the growing season, and N is the number of input variables in the ANNs, i.e., 10 (Fig. 2). The first row in the input matrix contained the values recorded for each input variable during the first one-minute interval, the second row contained the values recorded during the second interval, and so on. The output vector, on the other hand, contained the average canopy temperatures measured by an IRT group at each one-minute interval. Data contained by the training datasets obtained in this way were normalized using a Z-score normalization to account for the differences in the magnitudes of input variables. The percentages of data in these datasets allocated for the training, validation, and testing of ANNs were 70%, 15%, and 15%, respectively.



Fig. 2. Arrangement of elements in ANNs used in this study. Elements in the input, hidden, and output layers allocate 10, 12, and 1 neurons, respectively. The number of input neurons correspond to the number of variables that were considered relevant for the estimation of average canopy temperatures estimated by a given IRT group mounted on the center pivot. The only output neuron in the ANN allocates such average temperatures.

Results

Time series of average crop canopy temperatures estimated by ANNs and measured by IRT groups mounted on the center pivot are displayed for the first and second case studies in Fig. 3 and Fig. 4, respectively. On July 12, the scan started at 11.3 h at an angle of 227°. The center pivot then advanced in a counter-clockwise direction through the SSE side of the field and entered the NNW side at approximately 13 h. The scan was completed at 14.2 h when the pivot reached 248°. Since all IRT groups scanned experimental plots in the SSE side (where the highest irrigation level was assigned to all plots) before 13 h, measured canopy temperatures before this time tended to be smaller than temperatures obtained in the NNW side (where irrigation levels varied) after this time (Fig. 3). Nevertheless, ANNs were capable of approximating the oscillating pattern displayed by measured canopy temperatures through the scan (Fig. 3), with a Root Mean Squared Error (RMSE) that ranged from 1.04 °C to 2.49 °C (Table 1). To assess the impact of using ANNs for irrigation management, their estimated canopy temperatures were used by the iCWSI and hybrid methods to recalculate the prescriptions of experimental plots using these methods. No difference was found between the prescription map obtained with canopy temperatures measured by IRTs. Hence, the accuracy of all ANNs tested in the first case study can be deemed as satisfactory.

Regarding the second case study, the scan started on July 24 at 11 h at an angle of 52°. The center pivot then advanced in a counter-clockwise direction through the NNW side of the field and entered the SSE side at approximately 12.5 h. The scan was completed at 13.7 h when the pivot arrived at 68°. Similar to the first case study, measured canopy temperatures tended to be smaller as the center pivot advanced through the SSE side of the field, i.e., after 12.5 h. As in the first case, ANNs were capable of approximating the oscillating pattern displayed by canopy temperatures through the scan (Fig. 4), with a RMSE that ranged from 2.14 °C to 2.77 °C (Table 2). When comparing the prescription maps obtained with canopy temperatures estimated by ANNs and canopy temperatures measured by IRTs, only one plot (out of 26 assigned either the iCWSI or hybrid methods) was assigned a different prescription (Fig. 5). Therefore, the accuracy of all ANNs tested in the second case study can be also deemed as satisfactory.



Fig. 3. Time series of canopy temperatures measured by IRTs and estimated by ANNs trained to forecast the average temperatures obtained by IRT groups a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6 during July 12 (DOY 193). IRT group 1 consists of the two IRTs closest to the pivot point and IRT group 6 consists of the two IRTs farthest from the pivot point. ANNs were trained using data collected during the first three scans that took place on June 26, July 7, and July 11.





Fig. 4. Time series of canopy temperatures measured by IRTs and estimated by ANNs trained to forecast the average temperatures obtained by IRT groups a) 1, b) 2, c) 3, d) 4, e) 5, and f) 6 during July 24, 2017 (DOY 205). ANNs were trained using data collected during the first six scans that took place on June 26, July 7, July 11, July 12, July 17, and July 20.

Root Mean Squared Error	IRT	IRT	IRT	IRT	IRT	IRT
(RMSE)	Group 1	Group	Group	Group	Group	Group
		2	3	4	5	6
All irrigation levels	2.06	1.04	1.16	1.52	2.49	2.10
30% irrigation level	1.21	1.07	1.52	0.76	4.02	1.49
50% irrigation level	2.38	0.70	1.18	0.56	3.29	3.11
80% irrigation level	2.14	1.11	1.03	1.84	1.58	1.91

 Table 1. Root Mean Squared Error (RMSE) of ANNs used in the first case study to forecast

 average canopy temperatures measured by IRT groups during the scan performed on July 12

Table 2. Root Mean Squared Error (RMSE) of ANNs used in the second case study to forecast average canopy temperatures measured by IRT groups during the scan performed on July 24

Root Mean Squared Error (RMSE)	IRT Group 1	IRT Group 2	IRT Group 3	IRT Group 4	IRT Group 5	IRT Group 6
All irrigation levels	2.77	2.64	2.72	2.18	2.14	2.42
30% irrigation level	3.20	3.29	5.04	2.30	3.07	3.78
50% irrigation level	2.52	1.34	2.40	2.29	2.12	1.28
80% irrigation level	2.72	2.70	1.67	2.11	1.81	2.15



Fig. 5. Prescription maps generated using canopy temperatures a) measured by a network of wireless IRTs mounted on the center pivot and b) estimated by ANNs using data collected on July 24, 2017 (DOY 205). Prescriptions are displayed as percentages of a pre-specified maximum irrigation depth of 30.5 mm (1.2 in). Only one plot (p8) received a different prescription when using the canopy temperatures estimated by ANNs.

Conclusions

Machine learning is a data analysis technique that can be used to solve a wide range of problems, since it uses previous data to train computational machines that can make predictions on new data with no human intervention. This study analyzes the feasibility of using a popular machine learning technique to estimate canopy temperatures in a field irrigated by a Variable Rate Irrigation (VRI) center pivot system. Results indicate that the machine learning technique tested can predict canopy temperatures with a satisfactory accuracy. Machine learning technology can be useful to add redundancy to an Irrigation Scheduling Supervisory Control and Data Acquisiton System (ISSCADAS) patented by scientists from ARS (Bushland, Texas). The addition of machine learning capabilities to the ISSCADAS can assist users when poor visibility conditions prevent the correct estimation of canopy temperatures using the network of sensing systems incorporated by the ISSCADAS.

Acknowledgements

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-67021-24420. Work reported in this paper was also accomplished as part of a Cooperative Research and Development agreement between USDA-ARS and Valmont Industries, Inc., Valley, NE (Agreement No.: 58-3K95-0-1455-M).

References

- Andrade, M. A., Shaughnessy, S. A. O., and Evett, S. R. (2015). ARSmartPivot v . 1 Sensor based management software for center pivot irrigation systems. *In Proc. ASABE Annual International Meeting*, St. Joseph, MI.
- Andrade, M. A., Choi, C. Y., Lansey, K., Asce, A. M., and Jung, D. (2016). Enhanced Artificial Neural Networks Estimating Water Quality Constraints for the Optimal Water Distribution Systems Design. *Journal of Water Resources Planning and Management*, (April). https://doi.org/10.1061/(ASCE)WR.1943-5452
- Andrade, M. A., Shaughnessy, S. A. O., and Evett, S. R. (2017). ARSPIVOT, A sensor based Decision Support Tool for the Integrated irrigation Management of VRI Center Pivot Systems, Oak Ridge Institute for Science and Education Sponsored by USDA-ARS Email : Alejandro.Andrade@ars.usda.gov USDA-ARS USDA-ARS, 5–14.

- Beale, M., Hagan, M., and Demuth, H. (2011). *Neural Network Toolbox User's Guide*. Natick, MA: Mathworks.
- Evett, S. R., O'Shaughnessy, S. A., and Peters, R. T. (2014). 8924031 B1. United States of America.
- Farifteh, J., Van der Meer, F., Atzberger, C., and Carranza, E. J. M. (2007). Quantitative analysis of saltaffected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, 110(1), 59–78. https://doi.org/10.1016/j.rse.2007.02.005
- Ferreira, P. M., Faria, E. A., and Ruano, A. E. (2002). Neural network models in greenhouse air temperature prediction. *Neurocomputing*, 43(1–4), 51–75. <u>https://doi.org/10.1016/S0925-2312(01)00620-8</u>
- He, F., and Ma, C. (2010). Modeling greenhouse air humidity by means of artificial neural network and principal component analysis. *Computers and Electronics in Agriculture*, 71(SUPPL. 1). <u>https://doi.org/10.1016/j.compag.2009.07.011</u>
- Heaton, J. (2008). *Introduction to Neural Networks with Java* (2nd ed.). st. Luis, MO: Heaton Research, Inc.
- Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J. (1981). Canopy temperature as a Crop Water Stress Indicator. *Water Resources Research*, 17(4), 1133–1138. Retrieved from <u>http://onlinelibrary.wiley.com/doi/10.1029/WR017i004p01133/abstract</u>
- Kavdir, I. (2004). Discrimination of sunflower, weed and soil by artificial neural networks. *Computers and Electronics in Agriculture*, 44(2), 153–160. https://doi.org/10.1016/j.compag.2004.03.006
- Kaul, M., Hill, R. L., and Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. Agricultural Systems, 85(1), 1–18. https://doi.org/10.1016/j.agsy.2004.07.009
- O'Shaughnessy, S. A., Evett, S. R., Colaizzi, P. D., and Howell, T. A. (2013). Wireless Sensor Network Effectively Controls Center Pivot Irrigation of Sorghum. *Applied Engineering in Agriculture*, 29(6), 853–864. https://doi.org/10.13031/aea.29.9921
- O'Shaughnessy, S. A., Andrade, M. A., and Evett, S. R. (2017). Using an integrated crop water stress index for irrigation scheduling of two corn hybrids in a semi-arid region. *Irrigation Science*, *31*(5), 451–467. https://doi.org/10.1007/s00271-017-0552-x
- Pham, X., and Stack, M. (2018). How data analytics is transforming agriculture. *Business Horizons*, 61(1), 125–133. https://doi.org/10.1016/j.bushor.2017.09.011
- Schaap, M. G., Leij, F. J., and Van Genuchten, M. T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal*, 62(4), 847–855. <u>https://doi.org/10.2136/sssaj1998.03615995006200040001x</u>
- Yang, C., Prasher, S., Landry, J., Ramaswamy, H., and Ditommaso, A. (2000). Application of artificial neural networks in image recognition and classification of crop of weeds. *Canadian Agricultural Engineering*, 42(3), 147–152. https://doi.org/10.1007/s00417-014-2568-8