https://doi.org/10.17221/86/2021-RAE

Comparison of the machine learning and AquaCrop models for quinoa crops

*Rossy Jackeline Chumbe-Llimpe, Stefany Dennis Silva-Paucar, Yvan Jesús García-Lopez**

Department Industrial Engineering, Faculty of Engineering, University of Lima, Lima, Peru

**Corresponding author: ygarcia@ulima.edu.pe*

Citation: Chumbe-Llimpe R.J., Silva-Paucar S.D., García-Lopez Y.J. (2023): Comparison of the machine learning and AquaCrop models for quinoa crops. Res. Agr. Eng., 69: 65–75.

Abstract: One of the main causes of having low crop efficiency in Peru is the poor management of water resources; which is why the main objective of this article is to estimate the amount of irrigation water required in quinoa crops through a comparison between the machine learning and AquaCrop models. For the development of this study, meteorological data from the province of Jauja and descriptive data of quinoa crops were processed and a simulation period was established from June to December 2020. From the simulation carried out, it was determined that the best model to predict the required irrigation water is the Adaptive Boosting (AdaBoost) model in which it was observed that the mean and standard deviation of the AdaBoost models (mean = 19.681 and SD = 4.665) behave similarly to AquaCrop (mean = 19.838 and SD = 5.04). In addition, the result of ANOVA was that the AdaBoost model has the best *P*-value indicator with a value of 0.962 and a smaller margin of error in relation to the mean absolute error (MAE) indicator with a value of 0.629. Likewise, it was identified that, for the simulation period of 190 days, 472.35 mm of water was required to carry out the irrigation process in red quinoa crops.

Keywords: AdaBoost; irrigation system; predictive analysis; statistical analysis; water management

Nowadays, agricultural processes are exposed to cost overruns due to their high dependence on climatic conditions, such as droughts, frosts, floods, and pests [Ministry of Agrarian Development and Irrigation (Midagri 2020)].

According to the Midagri (2020), 'it is necessary to develop constant adaptation processes', since Peru is considered a megadiverse country due to its high number of plant species, among them, quinoa.

Quinoa is an herbaceous plant, whose main characteristic is its adaptability to different agricultural conditions, one of them is temperature (a range from –8 to 30 °C) and it can be cultivated up to 3 800 m a.s.l. Likewise, it can be grown in alkaline or saline soils, it can cope with various conditions, such as water shortage or excess water, and it needs organic matter with a pH between 5.5 to 7 (Midagri 2020).

Since water is an important resource for the development of crops and is of limited availability in high Andean areas, it is important to know each type of crop's water demand, taking the depth of the roots and the soil type the crop is grown in into account (Brentan et al. 2017).

A market analysis carried out by Midagri (2020) mentions that the departments of Peru that have

https://doi.org/10.17221/86/2021-RAE

the highest participation in quinoa production are Puno, Junín, Ayacucho, Cuzco, and Arequipa.

Due to the fact that Peru has a high index of national and export requirements for this food, this article seeks to identify a prediction model for the amount of irrigation water that is adequate to increase the yield of quinoa crops through the application of machine learning techniques and simulation of the irrigation processes using AquaCrop software (version 6.1).

The software developed by FAO's Land and Water Division, 'AquaCrop', helps to simulate crop yields in response to the water management system, considering that factor as a limitation. For the execution of the model, it uses a small number of parameters, balancing simplicity and precision (Food and Agriculture Organization of the United Nations 2021).

Machine learning is a technique that involves the formulation of algorithms for the analysis of historical and current information that helps to predict and infer the behaviour of various factors in the agricultural process in order to understand, learn and know how to act in the face of various situations which may affect the efficiency of the crop (Hablemos del campo 2018). Likewise, each algorithm has different requirements, processes and modelling times to achieve the expected result (Kao and Venkatachalam 2021). For this study, Adaptive Boosting (AdaBoost), decision tree, random forest and neural network methods were considered.

Although there are a limited number of agronomy studies related to artificial intelligence, the objective of this study is to compare the predicted amount of irrigation water for quinoa crops established by the AquaCrop software with the results of the machine learning models.

The research hypothesizes that there is a machine learning model that can provide similar results to those obtained by the AquaCrop software.

MATERIAL AND METHODS

The quantitative-continuous methodology was used in the study since numerical data collected from quinoa crops and meteorological data from the province of Jauja were studied, which were analysed using statistical procedures.

The research process began with the collection of historical red quinoa crop data and meteorological data from Jauja, both being considered necessary input variables for the calculation of the water balance. Having reliable and historical information allows for the better estimation and prediction of environmental conditions (Aguilar Aguilar and Obando-Díaz 2020). Due to this, standardisation processes and cleaning the input data were carried out using Knime software (version 4.7.0).

Next, the calibrated data was uploaded into the software Orange (version 3.32) for simulation in different machine learning models and into the AquaCrop software in order to obtain the estimated amount of irrigation water for the quinoa crops. The results of the water balance of both software packages were ordered, analysed, and compared.

Study area. Figure 1 shows the province of Jauja which is located in the department of Junín, it has an approximate territorial extension of 3 749.1 km^2, an altitude of 3 389 m a.s.l., with the coordinates of 11°46'34''S and 75°29'48''W (INEI 2017).

According to the productive profile of the quinoa crops in the province of Jauja, during the period of August 2019 to July 2020, 984 hectares were harvested, which had a yield of 1.58 tons per hectare,

Figure 1. Map of Peru and the Jauja province

https://doi.org/10.17221/86/2021-RAE

which means 1 549.8 tons of quinoa were produced (Midagri 2020).

The province of Jauja has 19 288 agricultural producers, it also has 239 613 hectares of agricultural land, of which, 65.1% of the cultivated area is intended for sale, that is, 155 988 hectares (Midagri 2020).

Evapotranspiration (*ETo*). Evapotranspiration consists of two processes in which the crop loses water, by evaporation through the soil surface and by transpiration through the plant tissues.

The equation for calculating the *ETo* is as follows:

$$ETo = \frac{0.408\Delta\left(R_n - G\right) + \gamma \frac{900}{T + 273} u_2 \left(e_s - e_a\right)}{\Delta + \gamma\left(1 + 0.34 u_2\right)} \qquad (1)$$

where: ETo – reference evapotranspiration (mm·day^{-1}); R_n – net radiation at the crop surface (MJ·m^{-2}·day^{-1}); G – soil heat flux density (MJ·m^{-2}·day^{-1}); T – air temperature at a height of 2 m (°C); u_2 – wind speed at a height of 2 m (m·s^{-1}); e_s – saturation vapour pressure (kPa); e_a – actual vapour pressure (kPa); $e_s - e_a$ – saturation vapour pressure deficit (kPa); Δ – slope vapour pressure curve (kPa·°C^{-1}); γ – psychrometric constant (kPa·°C^{-1}).

For this investigation, CropWat software (version 8.0) was used to obtain the *ETo* by entering the climatic variables, such as the relative humidity, maximum temperature, minimum temperature, heliophany and wind speed (Food and Agriculture Organization of the United Nations 2021).

AquaCrop. The process shown in Figure 2 simulates the separation of the soil evaporation from the crop transpiration and biomass yield, allowing for the more realistic accounting view of a dynamic nature, including the effects of the water stress and the crop responses (Food and Agriculture Organization of the United Nations 2021).

AdaBoost. This algorithm assigns the same amount of weight to all the training data in order to have a learning phase, then updates these weights for more accurate prediction results, assigning a higher weight to data with incorrect predictions and a lower weight to the correct ones, in a way that gives greater importance to those that are wrong. This process is carried out until the selected num-

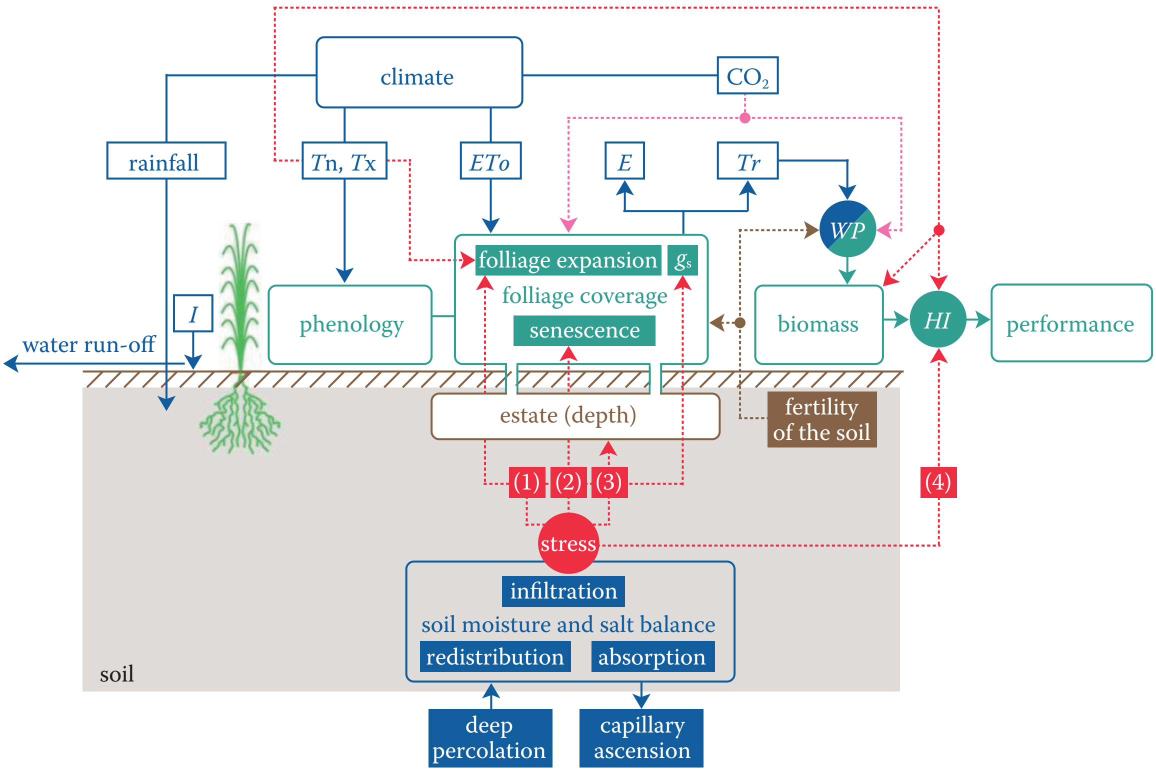

Figure 2. AquaCrop flowchart

Tn – air normal temperature; *Tx* – air maximum temperature; *ETo* – reference evapotranspiration; *E* – soil evaporation; *Tr* – actual canopy transpiration; *I* – irrigation; g_s – stomatal conductance; *WP* – water productivity coefficient; *HI* – harvest index; (1), (2), (3), (4) – feedbacks/feedforward from water stress

https://doi.org/10.17221/86/2021-RAE

ber of iterations or the desired error rate is reached (Wu et al. 2020).

The steps of the AdaBoost model for a regression problem can be expressed as follows:

(*i*) Initialise the weight distribution of the training samples, for $i = 1, 2, 3, \ldots M$

$$D_1 = \left(W_{11}, \ldots, W_{1i}, \ldots W_{1M}\right), W_{1i} = \frac{1}{M} \tag{2}$$

(*ii*) For k ($k = 1, 2, 3, \ldots K$), taking Dk as the training set of the weak learner $f_k(x)$ and calculating the following indicators:

– Maximum error:

$$E_k = \max\left|y_i - f_k\left(x_i\right)\right|, i = 1, 2, 3, \ldots M \tag{3}$$

– Relative error of each sample:

$$e_{ki} = \frac{\left[y_i - f_k\left(x_i\right)\right]^2}{E_k} \tag{4}$$

– Regression error rate:

$$e_k = \sum_{i=1}^{M} \omega_{ki}\, e_{ki} \tag{5}$$

– Weight of the weak learner $fk(x)$:

$$\alpha_k = \frac{e_k}{1 - e_k} \tag{6}$$

– Weight distribution of the samples is updated as:

$$\omega_{k+1,i} = \frac{\omega_{ki}}{z_k}\alpha_k^{1-e_{ki}} \tag{7}$$

where: z_k – normalising factor.

$$Z(k) = \sum_{i=1}^{M} \omega_{ki}\,\alpha_k^{1-e_{ki}} \tag{8}$$

(*iii*) The final strong learner is obtained as:

$$f(x) = \sum_{m=1}^{M}\left[\ln\left(\frac{1}{\alpha_m}\right)\right]g(x) \tag{9}$$

where: $g(x)$ – median of all $\alpha_m fm(x)$; $m = 1, 2, 3, \ldots M$ (Wu et al. 2020).

Neural network. The main function of the artificial neural network is to develop characteristics similar to those of the human brain, such as self-adaptability, self-organisation and tolerance to errors, based on algorithms and a large number of iterations to achieve its objective; which is to minimise or maximise the objective function (Quiñones Huatangari et al. 2020).

The architecture of the neural network is based on the structure of its elements such as: layers, nodes or interconnected neurons. The input layer receives the data that enters the system, the hidden layers are also called intermediate layers, they are in charge of executing various patterns and processing the input data to generate a result (Quiñones Huatangari et al. 2020).

The output of a neuron [= $f(n)$] is carried out as follows:

$$n = \sum_{j=1}^{R} \omega_j x_j + b \tag{10}$$

where: $x_1, x_2, \ldots x_R$ – input values; $\omega_1, \omega_2, \ldots \omega_R$ – neuron weights; b – bias value; $f(n)$ – activation function (Marroquin-Peralta et al. 2021).

Decision tree. It is a data mining technique with binary segmentation, in which the algorithm generates a tree where the branches represent the decisions and these generate successive rules that may arise from an assumed decision, in addition, the algorithm starts with a root node and then is divided into sub-nodes, where each group is mutually exclusive. The optimal tree will be the one with the least complexity, that is, the one with the lowest rate of poor qualification. Therefore, the impurity function is used, which is a measure that allows one to determine the quality of a node, which is denoted by $i(t)$. Although there are several measures of impurity used as criteria for the partition, one of the most used ones is related to the concept of entropy (Beltrán and Barbona 2021).

$$i(t) = \sum_{j=1}^{k} p \times \left(j \div t\right) \times \ln p \times \left(j \div t\right) \tag{11}$$

where: $j = 1, \ldots k$ – number of classes of the categorical response variable; $p \times (j \div t)$ – probability of the correct classification for class j at node t (Beltrán and Barbona 2021).

Random forest. It is an aggregation of several classification and regression trees, where each decision tree is created randomly. In addition, this

https://doi.org/10.17221/86/2021-RAE

method uses Breiman's idea of 'bagging', in which the trees take two-thirds of the data for training into account and combine the results in such a way that some errors are compensated for the others and a prediction is obtained, which generalises better (Georganos et al. 2019).

For the evaluation and qualification of the machine learning models related to the evapotranspiration variable, 'cross validation' was used as the sampling methodology, which is used to adjust and evaluate each candidate model in separate data sets so that the performance evaluation is impartial (Lei 2020).

Likewise, the 'test on train data' was used as the sampling methodology to evaluate the water balance in all the models.

RESULTS AND DISCUSSION

Table 1 shows the descriptive statistics of the parameters from 2017 to 2020, which are important for the application of the AquaCrop model. The parameters are: heliophany (*H*), maximum temperature (T_{max}), minimum temperature (T_{min}), relative humidity (*HR*), wind speed (*W*) and evapotranspiration (*ETo*).

In relation to the data in Table 1, one can see that there is a greater dispersion of data referring to the wind speed, since there is a standard deviation of 73.21, compared to the other variables.

Likewise, the kurtosis shows that all the variables are platykurtic, which means that they are scattered and their curve is flatter than a normal bell curve.

Table 2 shows the correlation coefficient of the meteorological variables to be studied. Those variables that have a coefficient greater than 0.4 and less than –0.4 are considered as having good correlation.

From Figure 3, it can be seen that the evapotranspiration variable has a strong correlation with the variables: relative humidity, maximum temperature and heliophany.

Next, the trend of the data for each variable in the four-year period with monthly graduation is shown in Figure 4.

Table 3 shows the annual average of the values according to the different meteorological variables. In relation to this, a positive trend is observed in the variables of temperature, heliophany and radiation. On the other hand, it can be observed that the relative humidity and wind speed have a negative trend in the recent period.

Table 1. Descriptive statistics of the parameters for 2017–2020

Statics description	*ETo*	*HR*	T_{max}	T_{min}	*H*	*W*
Average	3.18	81.35	19.850	4.500	6.06	171.29
Standard deviation	0.61	6.48	2.010	3.230	2.78	73.21
Kurtosis	0.04	–0.08	0.732	–0.118	–0.83	1.69
Range	4.19	36.00	14.800	15.800	11.30	518.00
Minimum	1.43	60.00	11.200	–5.000	0.00	0.00
Maximum	5.62	96.00	26.000	10.800	11.30	518.00
Confidence level (95%)	0.03	0.34	0.100	0.170	0.14	3.88

ETo – evapotranspiration; *HR* – relative humidity; T_{max} – maximum temperature; T_{min} – minimum temperature; *H* – heliophany; *W* – wind speed

Table 2. Correlation coefficient of the parameters for 2017–2020

Coefficient description	*ETo*	*H*	T_{max}	T_{min}	*HR*	*W*
ETo	1.000	–	–	–	–	–
H	0.716	1.00	–	–	–	–
T_{max}	0.744	0.60	1.00	–	–	–
T_{min}	–0.140	–0.61	–0.26	1.00	–	–
HR	–0.450	–0.29	–0.38	0.22	1.00	–
W	–0.070	0.12	–0.10	–0.19	0.16	1.00

ETo – evapotranspiration; *H* – heliophany; T_{max} – maximum temperature; T_{min} – minimum temperature; *HR* – relative humidity; *W* – wind speed

https://doi.org/10.17221/86/2021-RAE

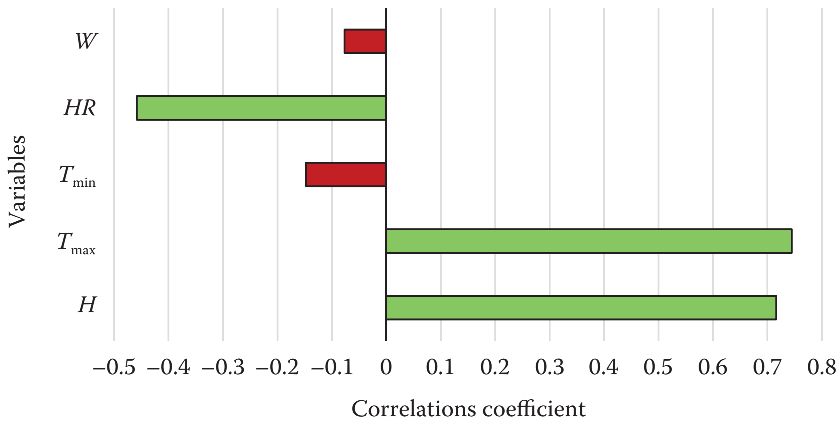

Figure 3. Correlation coefficient of the dependent and independent variables

W – wind speed; *HR* – relative humidity; T_{min} – minimum temperature; T_{max} – maximum temperature; *H* – heliophany

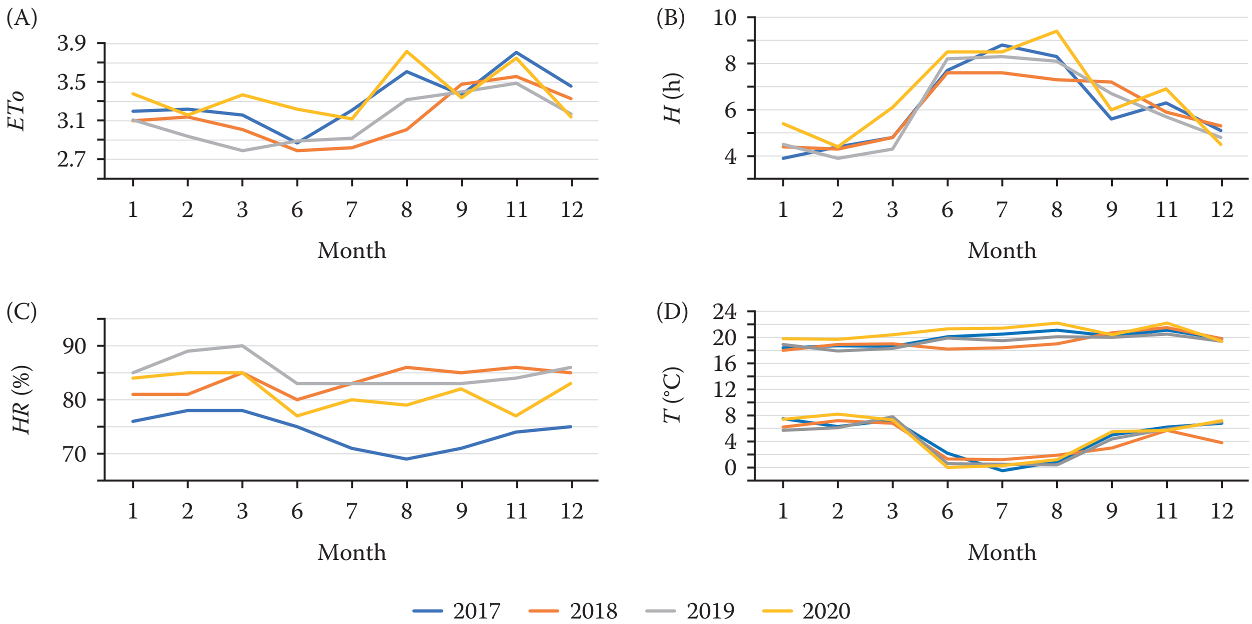

Figure 4. Trend of the independent variables – (A) *ETo*, (B) *H*, (C) *HR*, and (D) *T* for 2017–2020

ETo – evapotranspiration; *H* – heliophany; *HR* – relative humidity; *T* – temperature

Figure 5 shows the monthly water balance of evapotranspiration and precipitation from 2017 to 2020. This figure shows the variation of these parameters and their negative water balance for 32 out of a total of 45 months, that is, 71.11% of the study period. This graph helps to visualise the periods of water

Table 3. Annual average of meteorological variables for 2017–2020

Year	T_{min} (°C)	T_{max} (°C)	*HR* (%)	*W* (km·day^{-1})	*H*	Radiation (MJ·m^{-2})	*ETo* (mm)
2017	4.9	19.9	75	161	5.5	16.9	3.21
2018	4.1	19.3	85	192	6.1	17.7	3.09
2019	4.3	19.5	86	186	6.1	17.6	3.07
2020	4.8	20.7	81	171	6.6	18.6	3.40

ETo – evapotranspiration; *H* – heliophany; T_{max} – maximum temperature; T_{min} – minimum temperature; *HR* – relative humidity; *W* – wind speed

https://doi.org/10.17221/86/2021-RAE

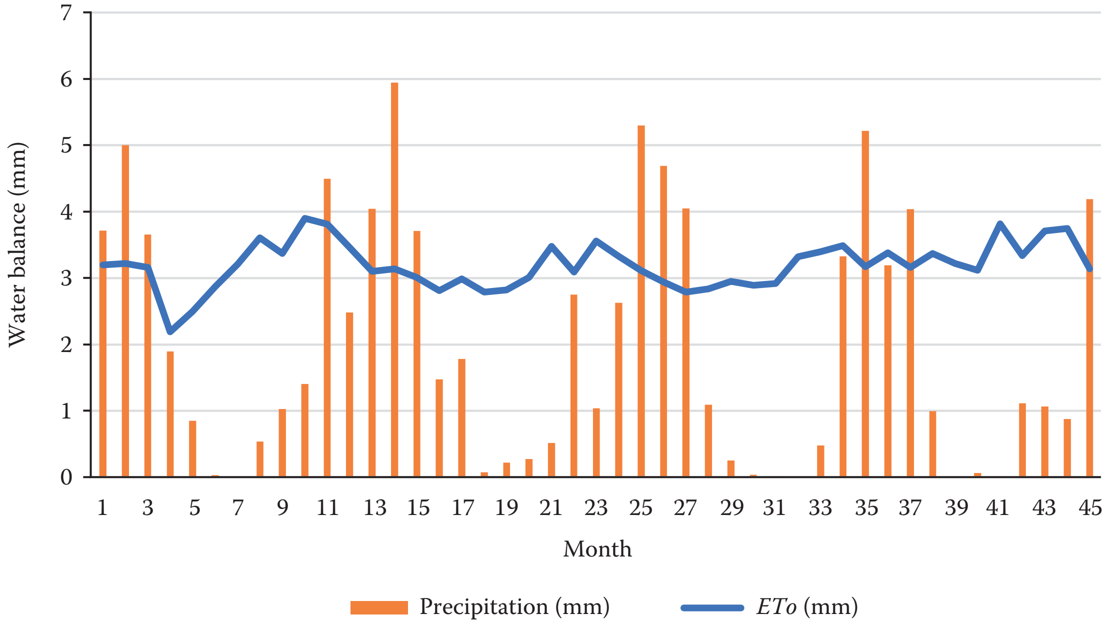

Figure 5. Monthly water balance for 2017–2020

ETo – evapotranspiration

deficit in the quinoa crops and, therefore, to know when, in which season, the irrigation water is needed in order to increase the crop yields.

A 2016 study carried out in the Bolivian Altiplano affirms that the yield of quinoa crops was not negatively affected in deficit irrigation conditions, despite the high climatic variability (Fajardo et al. 2016). However, it was shown that a water deficit exists in the quinoa harvest as a result of climate change, as shown in Figure 5, therefore, an optimal irrigation schedule is required considering the evapotranspiration variable. Similarly, a study carried out in 2015 in the Bolivian Altiplano had the same conclusions regarding the water deficit of quinoa crops (Alavi et al. 2015).

To work on the machine learning model, evapotranspiration was considered as the target and the simulation was carried out in the software Orange considering neural networks, AdaBoost, linear regression and random forest as the possible models.

The table below shows the results of mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination for the machine learning models, using the software Orange.

Table 4 shows that the neural network model has a higher determination coefficient compared to the other models, which means that it is a model that adjusts better to the real variables. This model has five input layers, three hidden layers and one output layer. In addition, Tanh was used as an activation function for the hidden layer and L-BFGS-B as a solver for the weight optimisation.

Water balance in the AquaCrop software. The software has four modules with input variables required for the simulation, which are: climate, cultivation, management practices, and soil. Within the climate module, data on the precipitation, evapotranspiration, maximum temperature, mini-

Table 4. Correlation coefficient and errors of the machine learning models

Models	MSE	RMSE	MAE	R^2
Neural network	0.060	0.245	0.187	0.839
AdaBoost	0.062	0.249	0.171	0.834
Random forest	0.066	0.257	0.186	0.822
Linear regression	0.077	0.278	0.223	0.792

MSE – mean square error; RMSE – root mean square error; MAE – mean absolute error; R^2 – coefficient of determination

https://doi.org/10.17221/86/2021-RAE

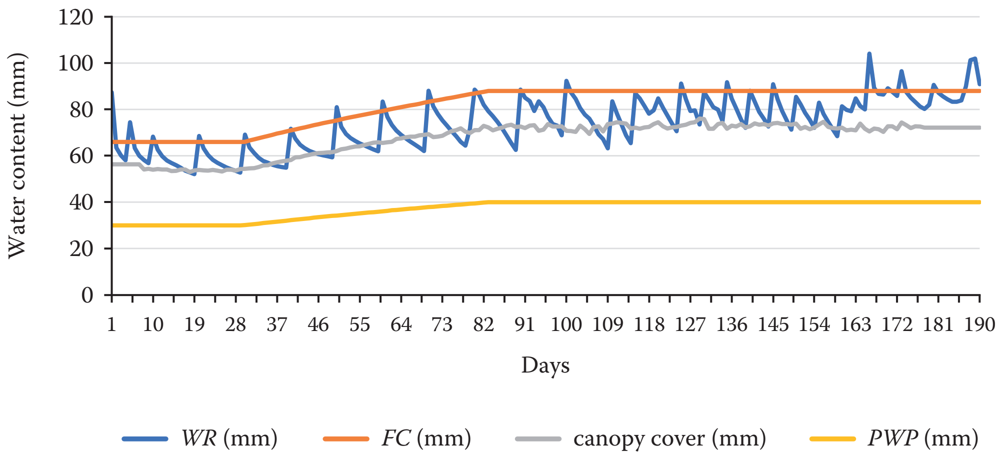

Figure 6. Water content in the root zone

WR – water content in the root zone; *FC* – field capacity; *PWP* – permanent wilting point

mum temperature and CO_2 were uploaded from June to December 2020. On the other hand, within the management module, sprinkling was selected as the irrigation method. In addition, within the soil module, the AC horizon (A – topsoil, C – parent material) was selected, which has a sandy loam texture. Finally, within the cultivation module, the red quinoa data were considered, whose sowing began on June 22, 2020 and ended on December 28, 2020.

As predicted by the software, the growing cycle was estimated as 7 days for the emergence, 125 days for the maximum canopy, 150 days for ageing and 190 days for maturity. Likewise, a maximum root depth of 0.4 m was observed, which was reached 83 days after sowing. To know the events of the deficit and excess water, the variables evapotranspiration, precipitation, evapotranspiration coefficient *(Kc)*, field capacity (*FC*) and the permanent wilting point (*PWP*) were measured.

Given this, the simulation carried out with the AquaCrop software showed the information for the said indicators helping to analyse the water situation of the crop, which can be seen in Figure 6. There is water stress in specific periods, since the water content in the root zone (*WR*) is below the canopy's threshold level, indicating the need for irrigation. In addition, periods with excess water can be observed due to the rainy seasons; this is evident in the peaks of the curve that are above the field capacity (*FC*).

To calculate the water balance, the precipitation and water infiltration were taken into account, which allowed us to know the irrigation events and the amount of water for each of them. The total amount of irrigation water during the evaluated period was 465 mm, taking the amount of precipitation as 168.5 mm, the amount of infiltration as 666.8 mm and the runoff as 2.9 mm into account.

On the other hand, Figure 7 shows the upward behaviour of the harvest index after the lag phase, which is the ratio of the biomass of the reserve organs and the total biomass (Food and Agriculture

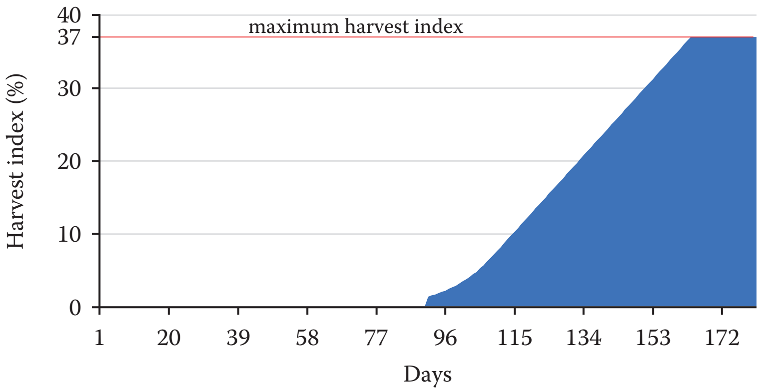

Figure 7. Harvest index of the quinoa crop

https://doi.org/10.17221/86/2021-RAE

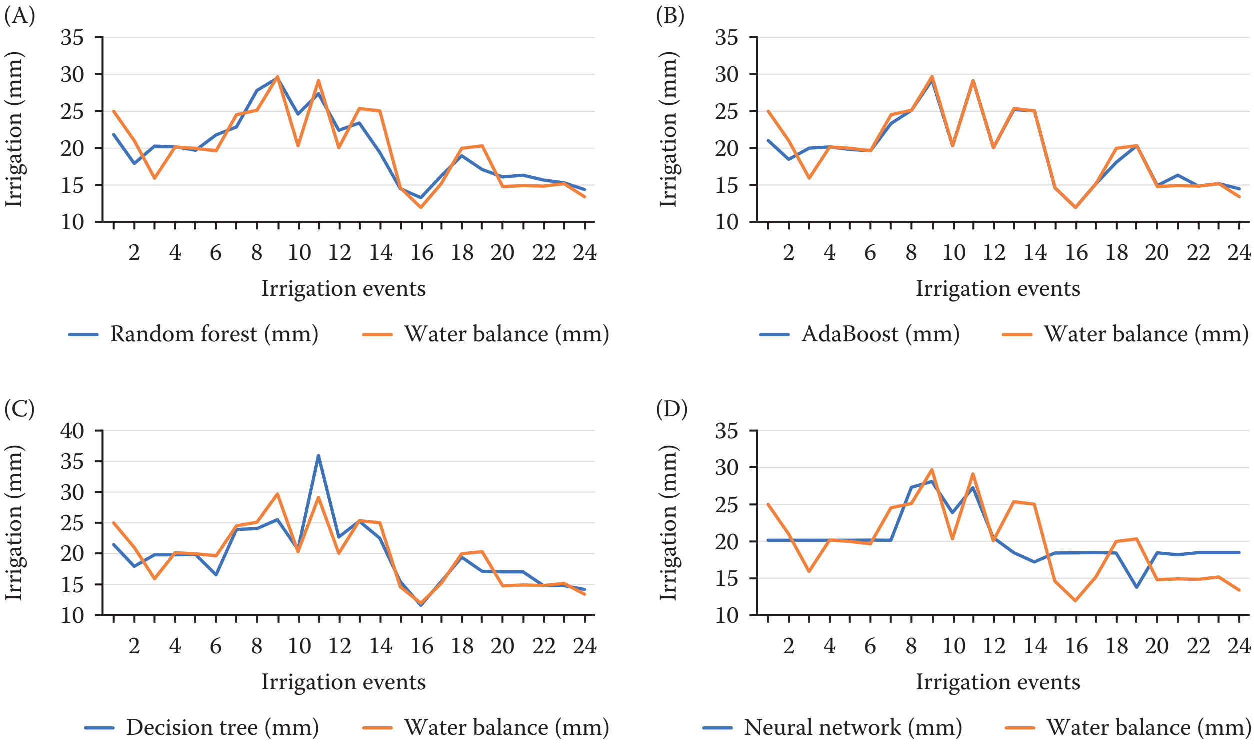

Figure 8. Comparison of machine learning models [(A) random forest, (B) AdaBoost, (C) decision tree, and (D) neural network] with AquaCrop

Organization of the United Nations 2021). The harvest index increases as the canopy cover increases. Once the canopy reaches a lower limit value, the harvest index reaches its final value, which is 37%, this value is within the range (0.06–0.87) which was calculated in a study carried out by the FAO in 2013 (Bazile et al. 2014).

In order to carry out a simulation that reflects the real behaviour of sowing quinoa in our study area, we considered analysing the data from the months of June to December, since the Integrated system of agricultural statistics of Midagri indicates that these months are part of the quinoa sowing calendar in Jauja (Midagri 2020).

Water balance in the machine learning models. Figure 8 shows the behaviour of each machine learning model and the water balance calculated with the variables: precipitation, crop coefficient, and evapotranspiration. It is evident that the AdaBoost model shows behaviour similar to the calculated water balance, unlike the other models.

Table 5 shows the values of the correlation coefficient, MSE, RMSE, and MAE calculated in the software Orange.

The neural network model has 13 input layers, five hidden layers and one output layer. In addition, Logistic was used as the activation function for the hidden layer and L-BFGS-B was used as the solver for the weight optimisation.

The result shows that the AdaBoost model has the highest correlation coefficient and the lowest error compared to the other models, which proves that it has the same behaviour as the water balance in the AquaCrop software.

Table 5. Correlation coefficient and errors of the machine learning models

Models	MSE	RMSE	MAE	R^2
AdaBoost	1.661	1.289	0.629	0.980
Random forest	10.821	3.290	2.338	0.872
Decision tree	11.652	3.413	2.223	0.863
Neural network	23.988	4.898	3.664	0.717

MSE – mean square error; RMSE – root mean square error; MAE – mean absolute error; R^2 – coefficient of determination

https://doi.org/10.17221/86/2021-RAE

Table 6. Results for the hypothesis according to the data of the machine learning models

Model	Mean	SD	Dispersion	Correlations*	*P*-value*	IC (95%)	MAE
AdaBoost	19.681	4.665	0.2320	0.980	0.962	(17.686; 21.676)	0.629
Random Forest	19.872	4.489	0.2210	0.872	0.938	(17.911; 21.833)	2.338
Decision tree	19.710	5.110	0.2538	0.863	0.929	(17.62; 21.79)	2.223
Neural Network	20.135	3.353	0.1630	0.717	0.811	(18.337; 21.894)	3.664
AquaCrop	19.838	5.040	0.2480	–	–	(17.99; 21.69)	–

* The results of the correlation and *P*-value evaluation were compared with the AquaCrop data; IC – index confidence; MAE – mean absolute error

According to the dispersion indicator shown in Table 6, the values of the models AdaBoost and random forest fit those of AquaCrop. In addition, evaluating the distribution of the data, it can be observed that the mean and standard deviation of the models have the same behaviour as that of AquaCrop. ANOVA was then performed, in which it was determined that the AdaBoost model has the best *P*-value indicator with a value of 0.962 and has a lower margin of error in relation to the MAE indicator. Therefore, the proposed hypothesis is accepted, which states that there is a machine learning model, in this case the AdaBoost model, which has behaviour similar to that of the AquaCrop software.

The study 'Validation of the AquaCrop model for different levels of fertility in the cultivation of quinoa in the Bolivian Altiplano' maintains that the AquaCrop software adequately simulates the life cycle of the plant, considering its simple management and its accessibility to the user that requires few input variables (Fajardo et al. 2016).

Nevertheless, in this investigation, to calculate the water balance in the AquaCrop software, it was necessary to include the cultivation, soil, and irrigation control as the input variables, in addition to the climatological data, unlike the calculation in the machine learning models, in which it was only necessary to input the climatological variables. Despite obtaining very similar results, it was shown that the effort made by the researcher in collecting the data necessary for the simulation in AquaCrop is greater than that of the machine learning model, since the FAO software requires more input data, to provide more details for all the parameters that affect the crop yield, therefore, the process becomes more complex, unlike the software Orange, which is more direct and simpler to process and analyse the data.

On the other hand, from the results obtained by the AquaCrop software and the machine learning models, it was determined that the AdaBoost model has a similar pattern to AquaCrop, since it is similar in percentage to the water balance of the AquaCrop of 99.61%, in comparison to the neural network model at 97.40%, random forest at 97.68% and classification tree at 98.15%.

However, these water balance results could change when adding input variables, such as the salinity and soil profiles, among others, that can cause water stress to quinoa crops, which can be considered in any future research as an opportunity to improve the precision for estimating the amount of irrigation water.

CONCLUSION

For this study, four machine learning models, which were neural networks, classification tree, random forest and AdaBoost, were compared to estimate the amount of irrigation water in quinoa crops. The precipitation, evapotranspiration and crop coefficient data were entered in each model, and the results were subsequently compared with the irrigation data obtained with the AquaCrop software.

The distribution of the data of the variables obtained for each model was evaluated, in which it was observed that the mean and standard deviation of the AdaBoost (mean = 19.681, SD = 4.665), random forest (mean = 19.872, SD = 4.489) and decision tree (mean = 19.71, SD = 5.11) models behave similarly to AquaCrop (mean = 19.838, SD = 5.04).

As a result of ANOVA, it turned out that the AdaBoost model has the best *P*-value indicator with a value of 0.962 and a smaller margin of error in relation to the MAE indicator with a value of 0.629.

Concluding, the initially proposed hypothesis is confirmed, which stated that there is a machine learning model, AdaBoost, that has similar behaviour to that of the AquaCrop software, with a total of 472.35 mm of irrigation water for a period of 190 days.

https://doi.org/10.17221/86/2021-RAE

On the other hand, it was shown that the AquaCrop software requires a greater number of input variables, such as climatological, crop control, soil and irrigation data to determine the water balance of a specific crop, while the software Orange that uses different machine learning models, such as AdaBoost, decision tree, neural network, random forest, among others, only needs weather data as the input variable, facilitating the processing and analysis of the information.

It should be noted that this work can be the basis for future research that considers more variables in the machine learning models in order to improve the accuracy of the water balance prediction.

REFERENCES

Aguilar Aguilar A.C., Obando-Díaz F.F. (2020): Aprendizaje automático para la predicción de calidad de agua potable. Ingeniare, 28: 47–62.

Alavi G., Diels J., Willems P., García M. (2015): Simulación de la producción de quinua en el altiplano boliviano con el modelo de Aquacrop con escenarios futuros generados por LARS-WG y QPM. Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales, 2: 7–13.

Bazile D., Bertero H.D., Nieto C. (2014): Estado del Arte de la Quinua en el Mundo en 2013. Santiago du Chili, FAO-CIRAD.

Beltrán C., Barbona I. (2021): Comparación del desempeño de Árboles de clasificación y Redes Neuronales en la clasificación politómica mediante simulación. Revista de Epistemología y Ciencias Humanas, 14: 2–11.

Brentan B.M., Luvizotto E., Herrera M., Izquierdo J., Pérez-García R. (2017): Hybrid regression model for near real-time urban water demand forecasting. Journal of Computational and Applied Mathematics, 309: 532–541.

Fajardo H., García M., Raes D., Van Gaelen H. (2016): Validación del modelo Aquacrop para diferentes niveles de fertilidad en el cultivo de quinua en el Altiplano Boliviano. Revista CINTEX, 21: 31–52.

Food and Agriculture Organization of the United Nations (2021): AquaCrop. Available at http://www.fao.org/aquacrop/overview/es/ (accessed July 22, 2021).

Georganos S., Grippa T., Niang Gadiaga A., Linard C., Lennert M., Vanhuysse S., Mboga N., Wolff E., Kalogirou S. (2019): Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International, 36: 121–136.

Hablemos del campo (2018): Machine Learning en la Agricultura. Available at https://www.hablemosdelcampo.com/machine-learning-en-la-agricultura/ (accessed Sept 17, 2021).

INEI (2017): Junin Compendio Estadistico 2017. Available at http://www.keneamazon.net/Documents/INEI/Junin.pdf (accessed July 22, 2021).

Kao Y.F., Venkatachalam R. (2021): Human and machine learning. Computational Economics, 57: 889–909.

Lei J. (2020): Cross-validation with confidence. Journal of the American Statistical Association, 115: 1978–1997.

Marroquin-Peralta J.M., Garcia-Lopez Y.J., Taquia-Gutierrez J.A. (2021): Machine learning: Comparison of algorithms for determining water quality in the Rímac river. Turkish Journal of Computer and Mathematics Education, 12: 552–572.

Ministry of Agrarian Development and Irrigation (Midagri) (2020): Análisis de Mercado 2015–2020. Quinua. Available at https://cdn.www.gob.pe/uploads/document/file/1479275/An%C3%A1lisis%20de%20Mercado%20-%20Quinua%202015%20-%202020.pdf (accessed Aug 16, 2021).

Wu T., Zhang W., Jiao X., Guo W., Hamoud Y.A. (2020): Comparison of five Boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. PLoS ONE, 15: e0235324.

Received: November 20, 2021
Accepted: November 11, 2022
Published online: May 2, 2023