# Predicting Daily Nutrient Water Consumption by Strawberry Plants in a Greenhouse Environment

Sathishkumar V E, Myeong-Bae Lee, Jong-Hyun Lim, Chang-Sun Shin, Chang-Woo Park, Yong Yun Cho
Dept. of Information and Communication Engineering, Sunchon National University

**Abstract**

Food consumption is growing worldwide every year owing to a growing population. Hence, the increasing population needs the production of sufficient and good quality food products. Strawberry is one of the world's most famous fruit. To obtain the highest strawberry output, we worked with three strawberry varieties supplied with three kinds of nutrient water in a greenhouse and with the outcome of the strawberry production, the highest yielding strawberry variety is detected. This Study uses the nutrient water consumed every day by the highest yielding strawberry variety. The atmospheric temperature, humidity and CO2 levels within the greenhouse are identified and used for the prediction, since the water consumption by any plant depends primarily on weather conditions. Machine learning techniques show successful outcomes in a multitude of issues including time series and regression issues. In this study, daily nutrient water consumption of strawberry plants is predicted using machine learning algorithms is proposed. Four Machine learning algorithms are used such as Linear Regression (LR), K nearest neighbour (KNN), Support Vector Machine with Radial Kernel (SVM) and Gradient Boosting Machine (GBM). Gradient Boosting System produces the best results.

## 1. Introduction

In the global economy, agriculture plays a critical role. Agrotechnology and precision farming, now also called digital farming, emerges as fresh scientific fields that use data-intensive methods to boost agricultural productivity while minimizing its effect on the environment. A multitude of different sensors gives the data produced in contemporary agriculture operations that enable a clearer knowledge of the operational environment (an interaction of dynamic crop, earth and weather conditions) and the operation itself (machine data), contributing to more precise and quicker decision making. In the present, technological advancement, such as the use of electronic devices and data transmission lead to revolutionary improvements in the agricultural working environment. These modifications require revised data from production systems and from agents and markets involved in production to provide decision-making knowledge for production as well as for strategic and managerial issues [1]. Together with large data technologies and high-performance computing, Machine Learning (ML) arises to create new possibilities to untangle, quantify and comprehend data-intensive processes in agricultural operational environments. Among other definitions, ML is the scientific field, which gives machines the ability to learn without being strictly programmed and inexpensive to make decisions.

Strawberry is one of the world's most famous fruits due to its production and nutrient gross value. Strawberry cultivation is seen as the key for the healthiest growing economy. Because of the elevated domestic strawberry demand for every season, the greenhouse cultivation is also rising. Since the strawberry plant has an entirely shallow root structure with branches reaching only about 6 inches long in sandy loam clay and because the flower blooms can be easily destroyed by summer frosts, proper irrigation is highly suggested. Water quality and good location for a strawberry bed yield a healthy berries plant for 3-5 years after that the plants start to decline, at which point the bed should be replanted. A bed of 100 plants provides about 100 quarters of berries, which is enough to provide plenty of healthy berries for a household of four.

## 2. Related Works

Commercial strawberry yield details in two districts of Norway are researched and compared with meteorological data [2]. The findings show that climatic factors are more crucial for output during flower induction and floral differentiation than conditions during flowering and ripening. The authors state that the regression method to predict strawberry yield may be more useful for strawberries production and marketing.

A yield prediction equation for the ' Strawberry Festival ' is created in [3] to enhance weekly forecasts using input variables obtained from floral counts and temperature data over two seasons in Florida.

Strawberries are cultivated from early December to late March in west-central Florida [4]. The prime harvest, usually from late February to mid-March, takes place at the beginning of the season and tends to take about 1 month. As the peak harvest advances and the temperature increases, fruit starts to shrink and the quantity of Soluble Substances Content (SSC) also reduces. The primary goal of this research is to assess whether peak harvest advancement leads to a temperature-independent drop in SSC. The findings show that increasing temperature in a subtropical system is a significant variable accountable for the late-season drop of SSC in strawberry fruit.

In 69 strawberry plantations, yield components are measured [5]. Fruit/hectare yield is positively correlated with crowndha count. The number of berries/trusses (inflorescence), average berry weight and fruit/truss weight are negatively associated with the number of trusses/crowns that emerge. Fruit/truss weight is positively associated with the number of berries/truss as well as the average weight of the berry. As the crown count stands for more than 50% of the yield variance, it is proposed that the future returns of strawberry plantations can be predicted from crown counts during the dormant period.

A modelling strategy for extracting the optimum yield curve for strawberry production in Florida that enhances farmers gain owing to competition from California and Mexico, and delicate market supply price responses are presented in [6]. The model incorporates Biological as well as financial limitations. Biological constraints for Florida output are designed on the premise that there are genetic and horticultural limitations for potential output enhancement, while financial limitations cater to cost modifications in response to supply.

## 3. Recorded Data and Description

In this experimentation, we use Greenhouse strawberry information from September 2018 to May 2019 (9 months). The greenhouse has three rows. Each row is supplied with different type pf Nutrient water. There are three beds in each row. In each bed, one variety of strawberry is cultivated. One thousand strawberries are planted in each bed. Each bed receives nutrient water regularly. Each bed has an inlet and outlet for nutrient water. By deducing Output nutrient water from Input nutrient water, the average nutrient water consumed by each strawberry plant is calculated.

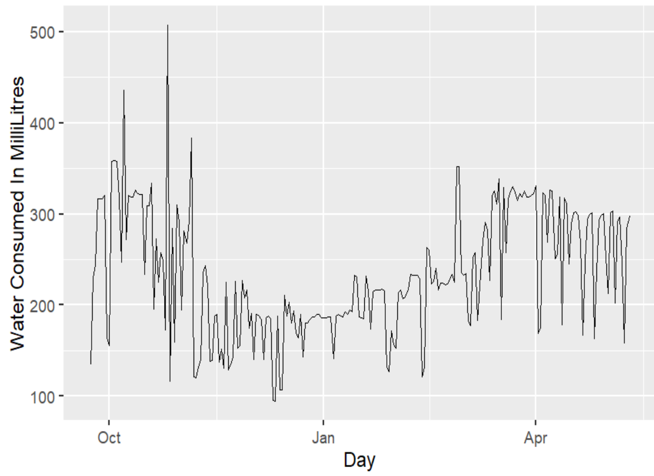
The readings of temperature, humidity, and CO2 are obtained for every minute using the sensors inside the greenhouse. Calculation of the maximum, minimum, average and gap (Maximum - Minimum) for temperature, humidity, and CO2 is done. These values are associated with the daily consumption of nutrient water by the strawberry plants and are used for the prediction.

| Parameters/Features | Abbreviation | Measurement |
|---|---|---|
| Date | Date | - |
| Water Consumed | Water | Milli Litres |
| Greenhouse Maximum Temperature | GHMaxTemp | °C |
| Greenhouse Minimum Temperature | GHMinTemp | °C |
| Greenhouse Average Temperature | GHAveTemp | °C |
| Greenhouse Temperature Gap | GHTempGap | °C |
| Greenhouse Maximum Humidity | GHMaxHum | % |
| Greenhouse Minimum Humidity | GHMinHum | % |
| Greenhouse Average Humidity | GHAveHum | % |
| Greenhouse humidity gap | GHHumGap | % |
| Greenhouse Maximum CO2 Value | GHMaxCO2 | ppm |
| Greenhouse Minimum CO2 Value | GHMinCO2 | ppm |
| Greenhouse Average CO2 Value | GHAveCO2 | ppm |
| Greenhouse CO2 Value Gap | GHCO2Gap | ppm |

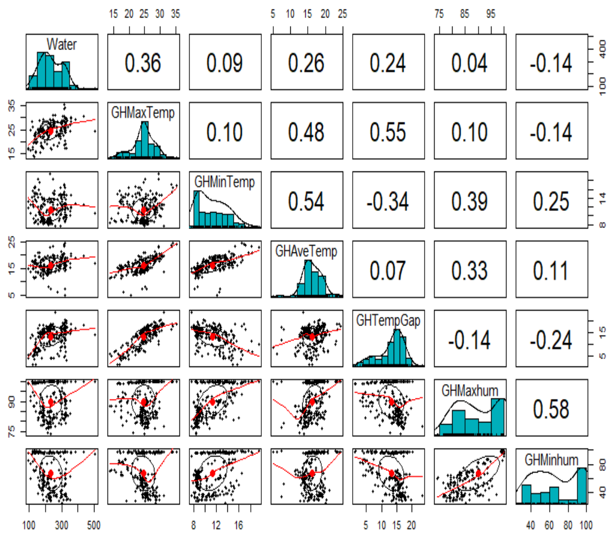
<Table 1> Data Variables and Description

Table 1 lists all variables or features or parameters and their respective abbreviation and measurement.

Fig 1 shows the nutrient water consumed by strawberry plants each day from September 2018 toMay2019.



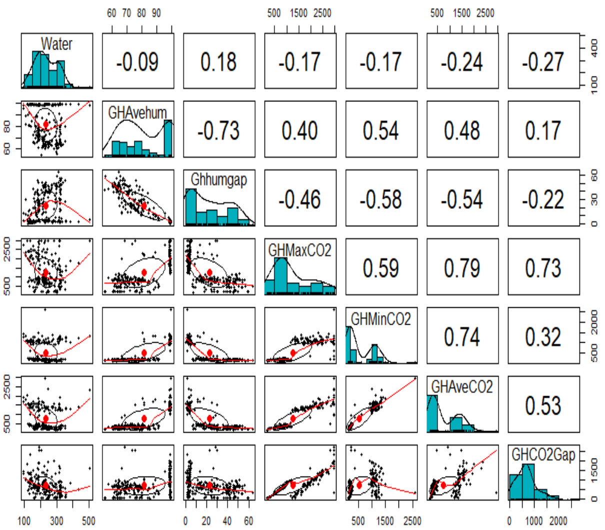

(Fig 1) Nutrient Water Consumption for the period of September 2018 to May 2019




(Fig 2) Pairs Plot. Relationship between Water with GHMaxTemp, GHMinTemp, GHAveTemp, GHTempGap, GHMaxHum, GHMinHum

Fig 2 and Fig 3 indicate pair plots illustrating the correlation between all variables with the Nutrient water consumed by the strawberry plant. This figure shows the bivariate scatter plots below the diagonal, histogram plots along the diagonal and the Kendall correlation above it. A correlation of 1 is a total positive correlation, -1 is the total negative correlation, and 0 is no correlation. For each pair, the linear regression line fits are displayed in red.

Fig 2 and Fig 3 shows a positive correlation between nutrient water consumption and temperature, GHMaxHum and GHHumGap. GHMinHum, GHAveHum, GHMaxCO2, GHMinCO2, GHAveCO2, and GHCO2Gap are negative correlations. Fig 2 Shows also a positive correlation between Water and GHMaxTemp (0.36), GHMinTemp, GHAveTemp, and GHTempGap. This shows nutrient water consumption increases with increase in temperature inside Greenhouse.




(Fig 3) Pairs Plot. Relationship between Water with GHAveHum, GHHumGap, GHMaxCO2, GHMinCO2, GHAveCO2 and GHCO2Gap

## 4. Model Selection

The dataset has 233 entries. The data is splitted into the train and the test. The training uses 75% of data and testing uses 25% of data.

<Table 2> Training and Testing Dataset

| Dataset | Number of observations |
|---|---|
| Training | 164 and 13 variables |
| Testing | 68 and 13 variables |

Four machine learning algorithms are used to predict nutrient water consumption: Linear Regression, K nearest neighbor (KNN), Support Vector Machine with Radial Kernel (SVM RBF) and Gradient Boosting Machine (GBM).

Regression algorithm efficiency is assessed using R2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) efficiency metrics. Table 3 shows the performance outcomes of the four regression models. The best-performing model is the model with the highest R2 value and the lowest RMSE, MAE and MAPE value.

| Model | Training | | | | Testing | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAE | MAPE | $R^2$ | RMSE | MAE | MAPE |
| LR | 0.38 | 55.68 | 41.87 | 19.50 | 0.09 | 60.97 | 48.13 | 22.19 |
| KNN | 0.54 | 48.17 | 34.24 | 15.60 | 0.19 | 57.42 | 41.79 | 19.21 |
| SVM | 0.56 | 46.95 | 30.58 | 13.95 | 0.37 | 50.40 | 36.53 | 17.38 |
| GBM | 0.63 | 42.87 | 28.40 | 12.85 | 0.45 | 47.19 | 34.79 | 16.43 |

<Table 3> Model Performances

As seen from Table 3. GBM has the highest R2 value of 0.63 and lowest RMSE (42,87), MAE (28.20) and MAPE (12.85) value in the train set. It also has the highest R2 of 0.45 and lowest RMSE (47.19), MAE (34.79) and MAPE (16.43) value in the test set compared to other regression models.

## 5. Conclusion

This research engaged in the development and comparison of four machine learning models to predict the daily consumption of nutrient water by strawberry plants with the Greenhouse Temperature, Humidity and CO2 data. Each model is trained using the regularly gathered data on nutrient water consumption and greenhouse data using four distinct data mining algorithms, namely linear regression, K nearest neighbor, Support Vector Machine with Radial Kernel and Gradient Boosting system. Relative to other regression algorithms, the performance of GBM is best. This research uses only 9 months' data. Since the data is small, Future work is to gather data on nutrient water consumption for 4 years and enhance the prediction performance.

### Acknowledgement

This work was carried out with the support of "Cooperative Research Program for Agriculture Science &Technology Development (Project No. PJ01188605)" Rural Development Administration, Republic of Korea and, this research was supported by IPET (Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries) through Advanced Production Technology Development Program, funded by MAFRA (Ministry of Agriculture, Food and Rural Affairs) (No. 315001-5)

## References

[1] Pivoto, Dieisson, et al. "Scientific development of smart farming technologies and their application in Brazil." *Information processing in agriculture* 5.1 (2018): 21-32.

[2] Døving, Aksel, and Finn Måge. "Prediction of strawberry fruit yield." *Acta Agriculturae Scandinavica, Section B-Plant Soil Science* 51.1 (2001): 35-42.

[3] MacKenzie, Steven J., and Craig K. Chandler. "A method to predict weekly strawberry fruit yields from extended season production systems." *Agronomy journal* 101.2 (2009): 278-287.

[4] MacKenzie, Steven J., et al. "The role of temperature in the late-season decline in soluble solids content of strawberry fruit in a subtropical production system." *HortScience* 46.11 (2011): 1562-1566.

[5] Mason, D. T., and N. Rath. "The relative importance of some yield components in East of Scotland strawberry plantations." Annals of applied biology 95.3 (1980): 399-408.

[6] Wu, Feng, Zhengfei Guan, and Vance Whitaker. "Optimizing yield distribution under biological and economic constraints: Florida strawberries as a model for perishable commodities." Agricultural Systems 141 (2015): 113-12