The International Society of Precision Agriculture presents the

15th International Conference on

Precision Agriculture

26–29 JUNE 2022

Minneapolis Marriott City Center | Minneapolis, Minnesota USA

Cabrera Dengra M[1], Ferraz Pueyo C[1], Pajuelo Madrigal V[1], Moreno Heras L[1], Inunciaga Leston G[2], Fortes R[1]

[1] HEMAV TECHNOLOGY, S.L. – Spain

[2] AB Azucarera Iberia, S.A. - Spain

USE OF MLP NEURAL NETWORKS FOR SUCROSE YIELD PREDICTION IN SUGAR BEET

15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract. *Sugar beet is one of the most technologically advanced agro-industries in Spain. In recent years, it has also led the digital transformation with the aim of maintaining the competitiveness of sugar beet both nationally and internationally. Among other lines, a very high potential has been identified in the determination of sucrose content through a combination of Artificial Intelligence and Remote Sensing. Artificial intelligence and machine learning application to agriculture could be an important tool to predict crop yields. The objetive of this study has been to create a predictive model of sucrose in sugar beet, with the aim of organizing logistics to increase the productivity by estimating the moment of maximum sucrose content and the validation of such model. For three years, satellite and drone spectral imaging, climatic and geographic information, as well as geo-referenced yield samples taken in the field were collected in all sugar beet production areas in Spain. These dates were analyzed and related to discriminate which were important for the model. For this, a Knowledge Discovery and Data Mining (KKD) process was carried out: collection of data, selection and cleaning, data mining and the generation of different models to reach the objective. Statistical knowledge was used to describe and understand the behaviour of the samples. Normalisations of the data and a clustering study with visualisation in Principal Component Analysis (PCA) was carried out to locate outliers at a multivariate level, and modelling the data with a multilayer perceptron (MLP) that is a fully connected class of feedforward artificial neural network (ANN). Data from 3,748 sucrose yield samples were predicted and related to real results, obtaining an accuracy measured by an* R^2 *of 0.9603 and a Mean Absolute Error (MAE) of 0.42. In its operational phase, increase in sucrose yield was validated in real fields. Supply chain managers used the sucrose prediction to determine the optimum harvest moment. The validation carried out in thirteen fields measured and increment of 8% of sucrose.*

Keywords.
Sugar beet, precision agriculture, neural network, deep learning, decision support systems,PCA, remote sensing, digitisation, data analysis, predictive analytics, data mining, drone, yield prediction, quality prediction.

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 15th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2018). Title of paper. In Proceedings of the 15th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Sugar beet (Beta vulgaris) is an important source of sugar for human consumption, because is one of just two crops which constitute the only important sources of sucrose. Sugar beet was first discovered as a potential sucrose source in 1802 in central Europe (Panella et al., 2014). Since then, it has been grown around the world as a primary sugar source alongside sugarcane. Sugar beet's provides nearly 30% of the world's annual sugar production and is a source for bioethanol and animal feed (Juliane C. Dohm, et al., 2014). Even though the sugar beet cropped area has been decreasing over recent decades, total production remains stable due to increasing yields. The highest average fresh root yield has been recorded in Spain (90 t/ha), despite it not being ranked among the world's 10 largest producing countries, which have yields ranging from 39 t/ha to 88 t/ha (FAOSTAT Database). In Spain in the 2013/14 campaign 26,605 ha were cultivated with a production of 2,135 Mt of winter-harvested-beet in the North Zone, and 8,662 ha were cultivated with a production of 749,502 Mt of summer-harvested-beet in the South Zone (MAP Spanish Agriculture Ministry - Department of agriculture).

In the framework of an industry that is undergoing important changes at the regulatory level, determining the sucrose content in sugar beet fields has been identified as very valuable for increase the yield. Therefore, the development of harvest prediction models has advanced, with the aim of predicting the sugar content of the crop before being harvested. There are currently only a few crop models available for simulating sugar beet growth and production. These models were developed based on either the empirical relationship between pre-harvested samples of sugar beet and final crop yield or the various plant growth processes involved at different growing stages (Vandendriessche and Van Ittersum, 1995). Empirical models include PIETER (Biemond et al., 1989; Smit et al., 1993), LUTIL (Spitters et al., 1989, 1990) and the model developed by Modig (1992). Process-based models include SUBGRO (Fick, 1971), SUBGOL (Hunt, 1974), SIMBEET (Lee, 1983), SUBEMO (Vandendriessche, 1989, 2000), SUCROS (Spitters et al., 1989), CERES-Beet (Leviel, 2000; Leviel et al., 2003), Broom's Barn (Qi et al., 2005), Green Lab (Vos et al., 2007), Pilote (Taky, 2008), and the model developed by Webb et al. (1997). Most of these models are based on environmental or physiological parameters or a combination of both.

Other models use spectral information collected using sensors carried on drones or satellites, to relate production to crop indices. Early in the sugar beet growing season, leaf area index (LAI) has been shown to be a good predictor of sugar beet yield (Clevers, 1997). Leaf area index is the projection of the leaf surface onto the soil as a proportion of the entire soil surface (Ross, 1981). Combining an estimate of LAI using remote sensing from aerial imagery or satellite within a crop growth model has been used to predict sugar beet yield (Clevers, 1997; Guerif and Duke, 1998; Hongo and Niwa, 2012). Remote imagery does not measure LAI directly but uses NDVI with red and near-infrared wavelengths as an estimator of LAI (Jordan, 1969). In our study, a further step has been taken to be able to integrate all the parameters described above, and to do so on a large scale through the combined use of remote sensing and environmental data, with the help of field sampling to adjust the calibration of the model. Neural Networks and Deep learning models have recently been used for crop yield prediction. You et al. (2017) used deep learning techniques such as convolutional neural networks and recurrent neural networks to predict soybean yield in the United States based on a sequence of remotely sensed images taken before the harvest. Their model outperformed traditional remote sensing-based methods by 15% in terms of Mean Absolute Percentage Error (MAPE). Russello (2018) used convolutional neural networks for crop yield prediction based on satellite images, using spatiotemporal features and outperformed other machine learning methods.

Our study set out to use Neural Networks and Deep learning to combine different types of parameters on commercial plots of sugar beet and compare the predicted results with the real ones. The main objectives of the development of this tool are to be able to identify the moment when the beet contains the highest sucrose value to be able to harvest on this time window maximizing total sucrose extracted from the same surface. In addition, this development will enable the evolution from predictive to prescriptive analytics in the sugar beet sucrose management.

At the beginning of the project, and as a tool for measuring confidence, it was established that the model must comply with a confidence ratio measured according to the parameter known as MAE (mean absolute error). It was established by geographical area (south, north-east, and north-west) and project phase. Thus, at the end of the 2020-2021 season (March 2021) the confidence in the tool must comprise an absolute sucrose error of 0.5 Tn/ha.

Material and methods

Study area

An action plan has been determined consisting of the collection of information of high punctual value using a large-scale root and foliar sampling methodology throughout the sugar beet areas in Spain: South, Northwest and Northeast.

The north of Spain is characterized by having extreme temperatures being colder and more humid than the south. That is the reason why in the south the sowing dates ranges from September to December and harvest is around June meanwhile in the northern area sowings are executed between January and April and harvest between October and March of the following year, due to lower temperatures.

The geographical distribution of the study is shown below. Colours represent the sugar mills where the sugar beet is processed: Guadalete (orange), Miranda (yellow), Toro (red) and La Bañeza (blue).

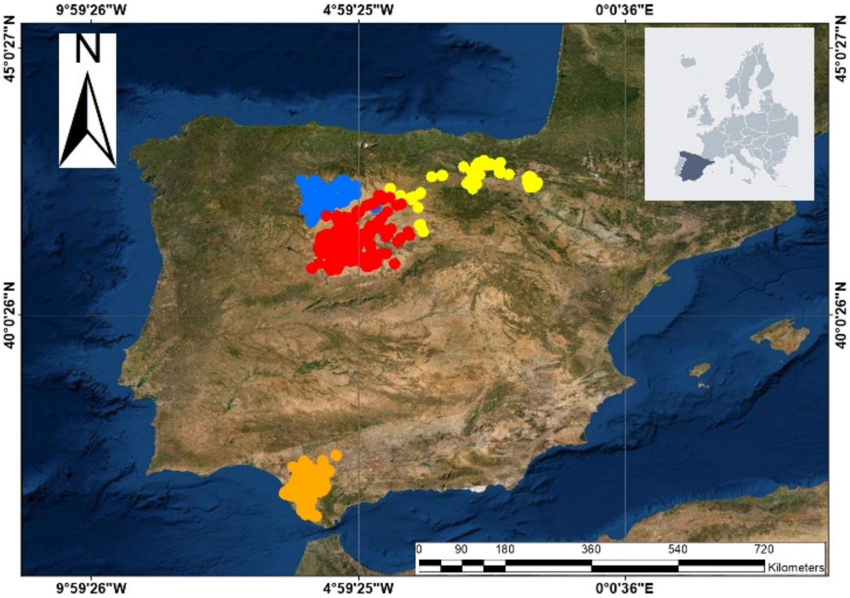

Figure 1. Field distribution by factory

In the selection of the plots to be sampled, the representativeness of the selection concerning the total contracting was considered. All the plots selected for the study must have the minimum data such as sowing date and type of irrigation.

Table 1. Summary of plots to be sampled by factory along three years.

Zone	Factories	Number of Fields	Area [Has]
South	Guadalete	88	1097
Northeast	Miranda	136	1086.48
Northwest	Toro	89	779.51
	La Bañeza	87	1095

Data acquisition

Field Sampling

Root and foliar samples from the selected plots were taken following an agronomically agreed methodology to ensure quality and homogeneity in the data collection. These samples have been analysed in laboratory obtaining quantitative parameters such as weight (both gross and net) and polarization. Other parameters have been gathered such as dry matter, reducing sugars mmo%, alpha-amino nitrogen mmol%, sodium mmol% or potassium mmol%.

Samples were taken using HEMAV LAYERS® sampling mobile app for geo-positioning.A protocol for sample collection and seasonality was designed to represent the sucrose value at the most relevant crop stages. In addition, a study of representativeness in the choice of plots was carried out before each sampling campaign.

Sensors

For this study, sensors were used for obtaining information from the vegetation. That is possible because vegetation has a low reflectivity in the visible spectrum, although with a peak in the green colour due to chlorophyll. Reflectivity is very high in the near-infrared due to the low energy absorption by plants in this band (Carmelo Alonso, 1999).

i) Micasense RedEdge

RedEdge is a multispectral camera specially designed for small drones and precision agriculture, environmental and forestry applications. It simultaneously captures images in five discrete spectral bands (R, G, B, RedEdge and nIR) listed in table 2.

Table 2: Micasense Bands

Band number	Band name	Wavelengths [nm]
Band 1	Blue (B)	480nm
Band 2	Green (G)	560nm
Band 3	Red (R)	670nm
Band 4	Red Edge	720 nm
Band 5	Near Infrared(nIR)	840nm

For this project, flights were planned at 120 m, which corresponds to a pixel size (GSD) of 8 cm/pixel.

ii) Sentinel 2

It is a multispectral satellite, part of the ESA Copernicus constellation, equipped with 13 bands (listed in table 3) distributed between the visible spectrum, near-infrared and shortwave infrared, which revisits the areas of interest every 5 days, maintaining the same viewing angles and thus allowing comparable information to be obtained.

Table 3: Sentinel 2 Bands used

Band number	Band name	Wavelengths [nm]	Bandwidth
Band 2	Blue	490 nm	10 m
Band 3	Green	560 nm	10 m
Band 4	Red	665 nm	10 m
Band 9	Short Wave Infrared (SWIR)	940 nm	60 m
Band 10	Short Wave Infrared (SWIR)	1375 nm	60 m
Band 11	Short Wave Infrared (SWIR)	1610 nm	20 m
Band 12	Short Wave Infrared (SWIR)	2190 nm	20 m

Band 1	Ultra blue (Coastal and Aerosol)	443 nm	60 m
Band 5	Visible and Near Infrared (VNIR)	705 nm	20 m
Band 6	Visible and Near Infrared (VNIR)	740 nm	20 m
Band 7	Visible and Near Infrared (VNIR)	783 nm	20 m
Band 8	Visible and Near Infrared (VNIR)	842 nm	10 m
Band 8a	Visible and Near Infrared (VNIR)	865 nm	20 m

The bands for the construction of the main vegetation indices are centred on 2, 3, 4 and 8, (E.G.Manrique, 1999) therefore, spatial resolution of 10m/pixel for most of the indices was obtained.

Climatology data

Climatic and agro-climatic variables provide information on how and to what the plant has been exposed from the date of planting to the date of sampling.

For the agro-climatic variables, GLDAS (Global Land Data Assimilation System) was used, which consists of providing information from satellite data and climatic stations, which were interpolated according to the terrain to obtain climatic data at each point of the terrain.

A weather API was used with a resolution of approximately 15-20 km depending on latitude. Source information came from information provided by weather stations and extrapolation taking into account the relief.

Lithological data

Soil type is considered an important condition for sugar beet production in Spain. To take this concept into account within the sucrose model, the geological map of the Iberian Peninsula at a scale of 1:1.000.000 has been used. This information has been downloaded in Shape format from the official website of the IGME (Geological and Mining Institute of Spain).

Variables

Different variables were analysed with the aim of adding to a mathematical model capable to predict sucrose at different harvest times, selecting a period when the beet contains the highest sucrose value before it starts decreasing.

For the elaboration of the predictive mathematical model, some dependent variables and one or several independent variables were required. All of them together with the methodology followed to ensure their quality are explained below.

As the objective of the project was the generation of a sucrose model, the main study variable of the project was sucrose, which acted as the dependent variable of the model.

Sucrose

Is the product of Polarisation (%) * Production (Tn/Ha). Therefore, it should be noted that this variable will always be influenced by these variables. All three were obtained from the field samples and analysed in the laboratory.

Production

Is given in gross weight to which a discount is applied to obtain the net weight of the sample. This weight corresponds to the weight of the roots present in the linear metres of sampling, already eliminating impurities such as soil.

The following conversion is used to extrapolate the weight of a sample to production in tonnes per average hectare of the plot.

$$(\text{Net weight [Kg]} * 2) / \text{linear m sampled} * 10000 / 1000 \quad (1)$$

Polarisation

Is given in percentage (%) of polarisation degrees. For the calculation in the laboratory, the regulations for the reception and analysis of beet are followed, which consists of cold digestion with lead sub-acetate.

Independent variables

Also known in a statistical context as regressors, represent potential reasons for variation. The values of the independent variables depend on the values of the dependent variables mentioned above. In the context of this project, they are the variables that influence in some way the sucrose content in sugar beet.

i)Spectral information

For the calculation of the spectral indices, the coordinates of each of the samples are taken into account to calculate them on the pixel in which the sample falls. From there, the reflectance of each of the bands from the wavelength of 443 (visible blue) to 2190nm (SWIR) is calculated to calculate the following plant indices (Ofer Beeri, 2004)

i.a) NDVI: Normalized Difference Vegetation Index.

The Normalized Difference Vegetation Index is used to estimate the quantity, quality and development of vegetation. It is calculated from the following bands:

$$NDVI = \frac{(NearIR - Red)}{(NearIR + Red)} \quad (2)$$

i.b) NDWI: Normalized Difference Water Index

The Normalised Difference Water Index can be used to identify water bodies and areas of high moisture saturation. In this way, we can use the index as a unit of measurement to determine water stress in vegetation and especially soil moisture saturation.

$$NDWI = \frac{860nm - 1240nm}{869nm + 1240\,nm} \quad (3)$$

i.c) NDRE: Normalized Difference RedEdge

This index is of interest when estimating the chlorophyll content of the plant, a factor that can be modified by stress conditions.

$$NDRE = \frac{(NearIR - RedEdge)}{(NearIR + RedEdge)} \quad (4)$$

i.d)GNDVI: Green Normalized Difference Vegetation Index

The GNDVI (Green Normalized Difference Vegetation Index) is an index of plant "greenness" or photosynthetic activity. It is one of the most widely used vegetation indices to determine water and nitrogen uptake in the crop canopy.

$$GNDVI = \frac{(NearIR - Green)}{(NearIR + Green)} \quad (5)$$

i.e)CCCI: Canopy Chlorophyll Content Index

Canopy Chlorophyll Content Index in relative content.

$$CCCI = \frac{\frac{(NearIR - RedEdge)}{(NearIR + RedEdge)}}{\frac{(NearIR - Red)}{(NearIR + Red)}} \quad (6)$$

i.f)TCARI: Transformed Chlorophyll Absorption Reflectance Index

The nitrogen index is obtained from the conversion of chlorophylls (TCARI INDEX) to nitrogen, direct ratio and internal HEMAV adjustment.

$$TCARI = \frac{3(700nm - 670nm) - 0.2(700nm - 550nm) * \frac{700nm}{670nm}}{(1+0.16) * \frac{800nm - 670nm}{800nm + 670nm + 0.16}} \tag{7}$$

Climatic variables

- Cumulative temperature which is the sum of the temperature averages from sowing date to sample.

- Accumulated precipitation (mm) is the sum of the daily accumulated precipitation in mm from the date of sowing to the sample.

- Accumulated degree days: can be defined as the heating or cooling requirements in degrees Celsius, necessary to reach the comfort zone, accumulated in a certain period of time. The following equation is used for the calculation.

$$GDD = [(Tmax\text{-}Tmin)/\ 2)] - Tb \tag{8}$$

where:
T max: maximum daily air temperature
T min: minimum air temperature
Tb: base temperature: the minimum temperature required for a crop to develop, it is dependent on the variety of sugar beet. For this case study, 3° has been taken as the base temperature.

This variable takes into account the difference between daytime and nighttime temperatures by taking the maximum daytime temperature and the minimum temperature as factors.

- Accumulated wind: Variable of how much wind the area has been subjected to. And in this case the sample.

- Maximum accumulated temperature: This is the accumulation of the maximum temperatures since sowing.

- Minimum accumulated temperature. This is the accumulation of the minimum temperatures since sowing.

- Maximum accumulated UV radiation.

Plot information variables

Plot data where the samples were taken such as irrigation type (rainfed or irrigated in the south), the sowing date, the previous crop and the plot identifier.

Sampling data: Analysis date, sample date, coordinates (latitude and longitude of the sample), number of beets per sample, linear metres of the sample.

Parcel data where the sample is taken: TC (type of irrigation), Farm Identifier (Identifier of the parcel) Contract, Date of sowing.

Generation of the data frame

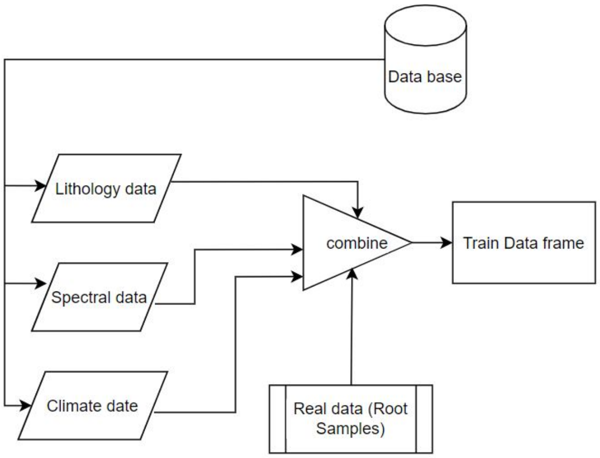

Figure 2:Workflow for dataset preparation

This first section of the study will cover the union of the data to a first exploration (as shown in Figure 2). The starting inputs are the information of the laboratory samples, their coordinates, plot information, layers codeand the information of all the independent variables that includes lithology, spectral and climate data.

Transformation of variables

The variables described include both continuous and categorical variables. The later are transformed as necessary. For example, the product of two variables may generate a new variable that provides additional information.

However, categorical variables are also available where several transformations are necessary to ensure that the results obtained for the variable in question are correct and interpretable. In this case, the original variable cannot be introduced into the model but, if the variable has n categories, each category must contain only 0 or 1 value.

Normalisation of variables:

For the optimisation of the model algorithms, it is necessary to normalise the input variables or, in other words, compress or extend the values of the variable so that they are all in a defined range and, therefore, can be comparable. This is especially important for those variables whose normal values are much larger than the rest. For example, the cumulative temperature ranges from 2000-4000ºC while the spectral indices range from 0-1.

The final objective is therefore to make all variables normalise within defined limits, which will be 0-1 (Sadaf Hossein Javaheri, 2014). For this purpose, the scaling of variables or also called Min Max Scaler is used.

$$\text{X normalized} = \frac{X - X\,min}{X\,\max - X\,min} \tag{9}$$

Where:
X is the original value without normalisation
X min: minimum value of the dataset for that variable
X Max: maximum value of the dataset for that variable

Process KDD (Knowledge Discovery and Data Mining)

Figure 3 shows the process of evaluating the data to arrive at the final model:

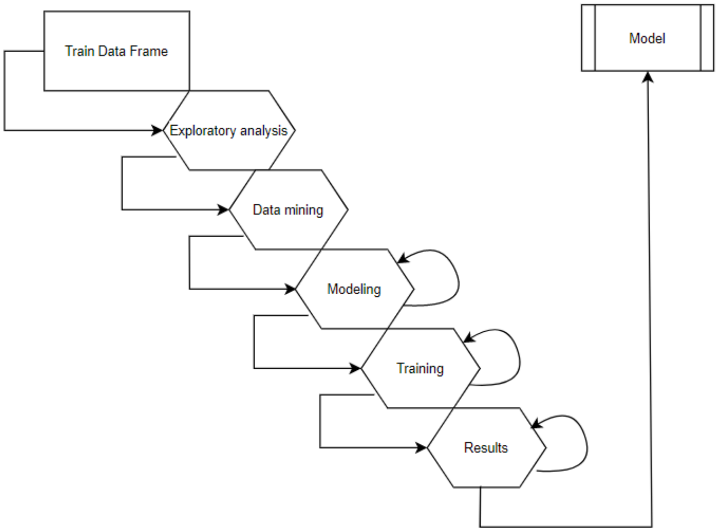

Figure 3 process KDD

Exploratory analysis

Univariate and multivariate are used to see the behaviour of the data for detecting outliers. Correlation matrices and histogram outliers are used to identify them.

In addition, we evaluate for the variable to be predicted, which of the independent variables has the highest correlation using Sklearn's SelectKBest library.

Data mining

Data mining was performed with unsupervised learning techniques. For the visualisation of the data, dimensions are reduced with principal components to be able to help us in the joint multidimensional visualisation of the dataset, and for the detection of these outliers, dendrograms are applied using parametric clustering. For the validations of the data considered outliers, it is used silhouette coefficient (Camilo Ordoñez, et al., 2020). It indicates the quality of groupings.

In this study, there are 58 variables in total and these techniques help us to evaluate the behaviour and purification of the variables that will be applied to predict sucrose.

Modelling

In the modelling of the data, the global set is divided into a train (80% of the data) and a test (20% of the data). This segmentation is random, but rather the histogram of the variable to be predicted is stratified, so that the test and train comprise a similar histogram and ensure values in all value ranges for train and test.

Training

In training, a neural network is generated. This is a supervised learning algorithm that learns a function $f(\cdot) : R^m \rightarrow R^o$ by training on a data set, where m is the number of dimensions for the input and is the number of dimensions for the output. Given a set of features and a target, it can learn a non-linear function approximator for classification or regression. It is different from logistic regression, in that between the input layer and the output layer, there may be one or more non-linear layers, called hidden layers.

Given a set of features and a target, it can learn a non-linear function approximator for classification or regression. It is different from logistic regression, in that between the input layer and the output layer, there may be one or more non-linear layers, called hidden layers.

Results

After the exploratory analysis, Figure 4 shows the different production trends for each of the mills. “La Bañeza” and “Miranda” mills were found to have a non-normal distribution.

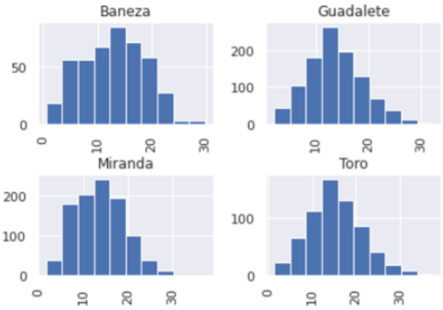

Figure 4: Sucrose distribution per factory

A population survey was carried out to identify possible candidates for outliers. These are carried out by mill in order not to mask outliers between different geographical areas. Polarisation and production variables are studied, as sucrose is a product of both.

Once the outliers of the dependent variables were detected, the multivariate analysis was performed by a heat map (Figure 5). This is helpful to determine which variables are the most relevant to the variable to be predicted, sucrose, and to detect variables that are homogeneous with each other.

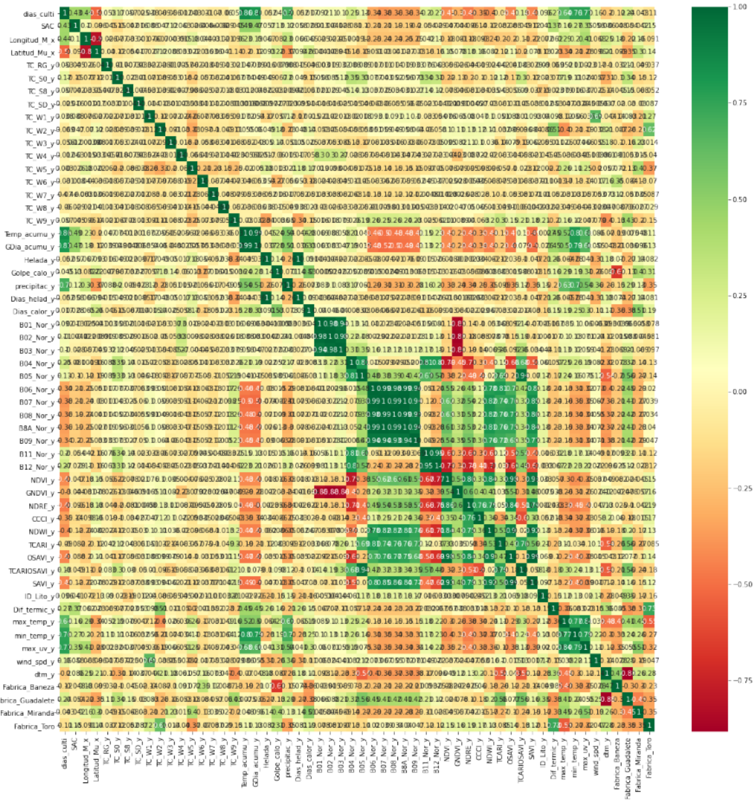

Figure 5: Multivariate analysis by heat map. Green colours mean more relationships between variables.

To assess whether outliers still exist, an unsupervised learning technique PCA (Principal Component Analysis) was applied. It was analysed how many PC according to the percentage variance are selected to describe the dataset. After carrying out the analysis, it was seen that one PC described 90% of the total dataset, and with three PC the 94% of the dataset are described.

This was the selected threshold on the PCA. Three dimensions PCA graph with cluster labels was executed . With this, 139 gauges create a set that is not able to cluster correctly, considering them as outliers.

Clusters and their dendrogram visualization were used for the location of the points that could be removed as outliers. The dendrogram is presented using the full method because it is quite robust and helps in the visualization of the different clusters. Figure 8**6** Figure 8 shows in red colour the level of values needed to adjust the dendrogram with a result of 6000 nodes. Silhouette coefficient methodology (Rousseeuw, 1987) also was used to validate these results obtaining a value of 0.6, enough to validate it, since values close to 1 are sought

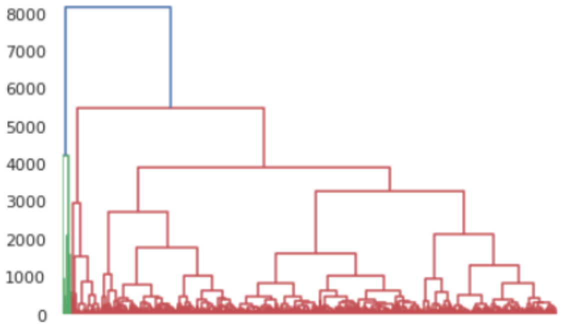

Figure 86: Completed dendrogram

Once the dataset was cleaned, it was normalised and the "Train" and "Test" datasets are prepared. This is done by stratifying the histogram (normalizing the bin size) to have the same histogram for test and train.

Following stetp after the dataset was prepared was the neural network generation with Keras. The proposed neural network was a multi-layer perceptron (MLP) with an inverse pyramidal structure. That is, we started with 500 neurons and go down to 1. The RELU activations helped to control the negative values of the weights and the last activation was the one that determined the regression value. We had to add L1 and L2 regularizers because otherwise, the model would have a severe overfit, i.e. the validation loss would worsen while the training loss would continue to improve.

Quality of the model train is presented in the following figure:

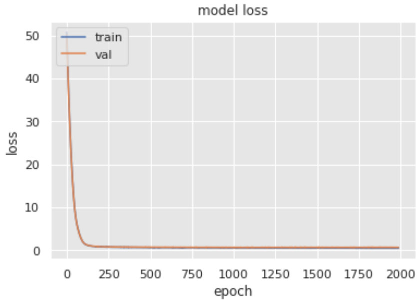

Figure 97: Quality of the model train

As shown in figure 9, 2,000 iterations were carried out, but Figure 9Figure 97 shows that at 250 iterations, the trainee stabilises. It is needed to highlight as positive that the test line follows the train line.

Training result obtained was R^2 0.9603, MAE 0.42 from 3,748 valid samples. In figure 10,Figure 108 left plot, it can be seen the value of sucrose from the laboratory in green vs predicted sucrose in red. Also, on the right plot, the correlation between these two variables in the right plot are shown.

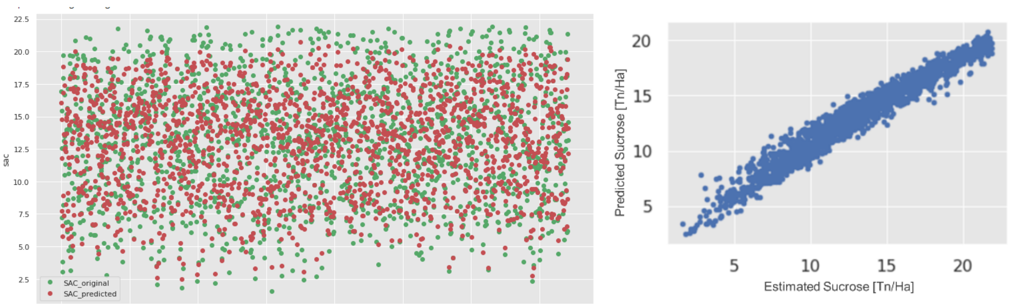

Figure 108: Original sucrose vs predicted sucrose

As the aim is to create a sucrose model, Figure 11 Figure 119shows the sucrose collected at the fields, already refined, on the growth degree days (GDD) axis.

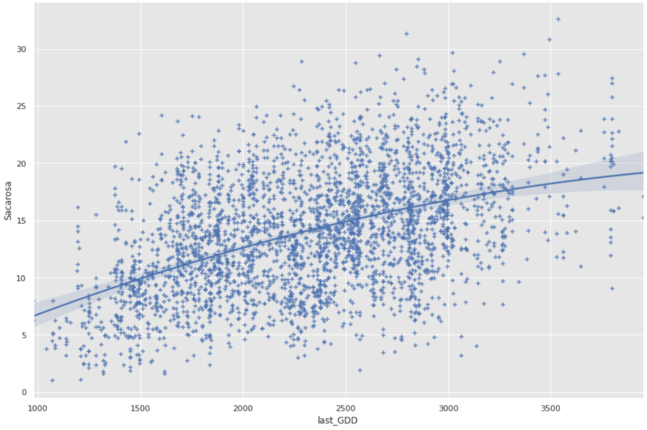

Figure 119: Sucrose [Tn/Ha] vs GDD

For the variables to be related, an exploration was made of the information that could be available, mainly from agriculture, sensors and climatology. In total, 55 variables were selected as candidates. Table 4: Samples taken and Outliers per factoryTable 4 represents the initial dataset debugging phases.

Table 4: Samples taken and Outliers per factory

Mill	Initial Samples	Final samples	Polarisation Outliers	Weight Outliers	Spectral Outliers	Global PCA Outliers
Guadalete	1374	1100	80	64	131	
Bañeza	877	769	12	6	20	
Miranda	1287	1039	53	25	170	
Toro	1038	840	43	39	47	
Global	3150	3748	188	134	368	139

As explained above, the variable we wanted to predict is a product of the polarisation and weight variables, which is why these variables were analysed to detect outliers. In order to do so, the same approach and modelling was performed but with polarization and production (net weight) instead of sucrose. This approach helped in finding polarization and production outliers.

Globally, and in view of the geographical differences, an analysis was made at mill level. The eliminated samples are reflected in the table above.

In total, 188 samples of polarisation (6% of the total sample) and 134 of the production variable (4.2% of the total sample) were removed. Despite not being a high percentage, it is considered that further strategies can still be implemented to save the maximum number of samples.

In the purification phase, 139 outliers were dropped after performing a clustering exploration with the whole global set visualised in a graph represented in PCA.

The results of the MLP model are shown in Table 5. It includes R^2 and Mean absolute error (MAE). It is the average, in absolute values, of the difference between sucrose prediction and the true value. The global result, considering samples from all mills is R^2 0.96 and MAE 0.42 which means the model can be wrong +-0.42 Tn sucrose/ha.

Table 5: Statistical results of the study.

Factory	Valid samples	R2	MAE
Guadalete	1100	0.9696	0.47
Miranda	1039	0.9691	0.36
Bañeza	769	0.9677	0.47
Toro	840	0.9596	0.43
Global	3748	0.96.03	0.42

After the validation presented here, the model was put into operation mode, providing updated information to farmers and supply chain managers in a weekly basis. For this reason, validations have been carried out to check its reliability and impact on the crop.

This validation was carried out on 13 randomly selected fields, taking samples at the harvest date proposed by conventional methods, and in the new harvest dates selected considering sucrose behaviour predicted by the sucrose model. As a result, a validated increase of 8% in production was achieved in these 13 plots following the recommendation of the model. It is estimated that the increase in production with these techniques could reach a theoretical potential of 20%.

Discussion

Several studies have tried to Model growth, development and yield of sugar beet using different tools for its parameterization. Models such as the Support System for Agrotechnology Transfer (DSSAT), developed by the United States Department of Agriculture in 2012, provides a common framework for a cropping system study used climatic and soil information for modelling. The model was successfully applied for predicting yield for six different sugar beet cultivars grown in North Dakota during the 2014 to 2016 growing seasons. Results could be applied for predicting sugar beet yield for different scenarios in regions with favorable environmental conditions for sugar beet production (Anar et al., 2019). Our model also uses climatic and edaphological parameters, but these parameters are not always capable to explaining specific events that may affect the development of the crop, such as the presence of pests or diseases, deficiencies in the application of irrigation or fertilizers, or incidents derived from the tillage before and during the crop cycle. Other models such as AquaCrop developed by FAO (The Food and Agriculture Organization) are useful for estimated crop yield response to water and fertilize (Raes et al., 2009b; Steduto et al., 2009). The overall performance of AquaCrop for simulating canopy cover, biomass, and final yield was accurate (R^2=0.924, R^2=0.957 and R^2=0.908) into different irrigation water allocations in the two main production areas of sugar beet in Spain (Garcia-Vila et al., 2019), but model must be accurately calibrated and validated taking field data manually, therefore, the size of the plant and its development should be evaluated more quickly and can even be automated to feed the models. Our model accesses information related to the size and development of the crop cycle through the evaluation of spectral information, which can be included into the modeling automatically. Streamlining the data collection to introduce in the models is necessary, since the degree of development of the crop and its crop health can influence the modeling as much as the soil or the climate does. For a more extensive and automated data collection, other studies face the modeling of the crop from its monitoring through spectral analysis, either with a drone or with a satellite image. Bu et al., 2015 analyzed the use of two optical sensors based on the normalized differential vegetative index (NDVI), both sensors were useful in providing sensor data that was related to yields from a series of harvest dates (R^2=0.59 $p<0.0001$). Sensor readings were most significantly related to yield within a site when root yield and recoverable sugar yield was related to Nitrogen rate. In another study, similar technology was used, but embedded in UAS

(Unmanned Aerial System), the advantage of these measurements is the possibility of taking mass georeferenced samples, which enables the development of prediction maps for large plots using geostatistical analysis with sufficient resolution to produce helpful information about crop management (Fortes et al., 2014). Chancia et al., 2021 used a multispectral image acquired by drones to relate the canopy cover of the sugar beet crop with its production level ($R^2 = 0.89$). This study demonstrates the potential for models using a combination of radiometric and canopy structure data obtained at early growth stages. The possibility of accessing this data in a massive way is possible through the satellite image. Large surfaces can be evaluated quickly, obtaining information about the level of spatial-temporal development of the crop, accessing the complete information of the crop cycle to introduce in the model. Beeri et al., 2004b evaluated reflectance indices using Landsat satellite images for 2002 and 2003 years, and the results suggested that these indices have potential to be used to predict the yield (R^2=0.88 $p<0.05$) and the sugar content (R^2=0.76 $p<0.05$) in sugar beet crop. However, to further improve the predictability, the study also proposed to integrate information from other data sources such as soil type, rainfall, air temperature, leaf evaporation and solar radiation during the growing season. Neuronal networks allow incorporating all the parameters used in the studies cited above, in addition to many others to find the relationships between them, and which parameters and values should be excluded or prioritized when modelling. Perhaps this great ability to incorporate and analyze parameters is responsible for reaching values close to 96% of R^2 obtained when we validate the estimates with the real data obtained in our study.

Conclusion

The study made it possible to make a prediction with an R^2 of 0.96, obtaining information that allowed the industry to increase sucrose by 8% just modifying the harvest date. Neural networks have been shown to be more accurate in predicting than other models. The automation in the data ingestion, and the use of neural network model allows, thanks to the variable weight metrics, enables speeding up the modeling tasks, scaling the processes by being able to evaluate large crop areas in a space-time manner.

Acknowledgements

This work has been possible thanks to a strong collaboration between AB Azucarera Iberia, S.A., leading sugar company in the Spanish market and part of the AB Sugar group, with a clear vision of an innovative opportunity and a potential competitive advantage, and the technical and scientifical capabilities of the AgriTech start-up HEMAV. This success is in our humble opinion an example of how big companies shall act as innovation leaders and tractors of their industries. It has not been an easy path but we thank for the vision and the hard work from all the AB Azucarera and HEMAV staff from the technical team to the Board of Directors.On the other hand, this project would not have been possible without the hard work of all the sugar beet farmers in Spain, serving this project as well as an example of technology applied to the technification and impact in rural areas; and the support from the Spanish sugar beet research center AIMCRA for the laboratory support without which all these data would not be acquired.

References

Anar, M.J.; Lina, Z.; Hoogenboomb, G.; Sheliab, V.; Batchelord, W.D.; Tebohe, J.M.; Ostliee, M.; Schatze, B.G.; Khanf, M. Modeling growth, development and yield of Sugarbeet using DSSAT. Agr. Syst. 2019, 169, 58–70.

Beeri, O.; Zhang, X.; Newcomb, T.; Carson, P.; Wagner, G. Using Landsat Images to Map the Quality and Quantity of Sugar Beet Yield; Sugarbeet Research and Education Board of Minnesota and North Dakota: Fargo, ND, USA, 2005; pp. 125-131.

Biemond, T., Greve, H.J., Schiphouwer, T., Verhage, A.J., 1989. PIEteR: Semi green-box produktiemodel suikerbieten. LU Wageningen, Vakgroep Agrarische Bedrijfseconomie, pp. 31.

Bu H., Sharma L. K., Denton A., Franzen D. W, 2015. Sugar Beet Yield and Quality prediction at multiple Harvest Dates Using Active-Optical Sensors. Agron. J. 108, 273–284. 10.2134/agronj2015.0268

Camilo Ordoñez, C., Méndez, C., Armando Ordoñez, H., & Armando Ordoñez, J. (2020). Evaluación e implementación de técnicas de. Iberian Journal of Information Systems and Technologies.

Carmelo Alonso, V. M. (1999). Determinación experimental de la firma espectral del a. VIII Congreso nacional de teledetección , 429-432.

Chancia R, Van Aardt J, Pethybridge S. J, Cross D, Henderson J. Predicting Table Beet Root Yield with Multispectral UAS Imagery. Remote. Sens. 13(11): 2180 (2021)

Clevers, J.G.P.W. 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. Remote Sens. Environ. 61:221–228. doi:10.1016/S0034-4257(97)00004-7

FAOSTAT Database. Available online: http://www.fao.org/faostat/en/#home (accessed on 15 July 2019).

Fick, G.W., 1971. Analysis and simulation of the growth of sugar beet (Beta vulgaris L.). Ph.D. thesis. University of California, Davis.

Fortes R, Prieto H, Millán S. Terrón JM, Blanco J. Campillo C, 2014. Using apparent electric conductivity and NDVI measurements for yield estimation of processing tomato crop. ASABE 57(3): 827-835.

Garcia-Vila M, Morillo-Velarde R and Fereres E, 2019. Modeling Sugar Beet Responses to Irrigation with AquaCrop for OptimizingWater Allocation. Water (IF3.103), Pub Date: 2019-09-14, DOI: 10.3390/w11091918.

Guerif, M., and C. Duke. 1998. Calibration of SUCROS emergence and early growth module for sugarbeet through optical remote sensing data assimilation. Eur. J. Agron. 9:127–136. doi:10.1016/S1161-0301(98)00031-8

Hongo, C., and K. Niwa. 2012. Yield prediction of sugar beet through combined use of satellite data and meteorological data. J. Agric. Sci. 4:251–261.

Hunt, W.F., 1974. Respiratory control and its prediction by a dynamic model of sugar beet growth. Ph.D. thesis. University of California, Davis.

Jig Han Jeong, J. P.-M.-H. (2016). Random Forests for Global and Regional Crop Yield Predictions. Journal.

Jordan, C.F. 1969. Derivation of leaf-area index from quality of light on the forest floor. Ecology 50:663–666. doi:10.2307/1936256

Juliane C. Dohm, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O. Himmelbauer , H. (s.f.). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris).

Lee, G.S., 1983. Conceptual Development of a sugarbeet crop growth model. Ph.D. thesis. Colorado State University, Fort Collins, Colorado.

Leviel, B., 2000. Evaluation des risques et maîtrise des flux d'azote au niveau d'uneparcelle agricole dans la plaine roumaine et bulgare. Application aux cultures demais, blé, colza et betterave. Institut National Polytechnique de Toulouse, France Ph.D. Thesis.

Leviel, B., Crivineanu, C., Gabrielle, B., 2003. CERES-Beet, a model for the production and environmental impact of sugar beet. In: A Proceedings of the Joint Colloquium on Sugar Beet Growing and Modelling, Sept. 12th, 2003. Lille, France.

Manrique.E.G. (1999). ÍNDICE DE VEGETACIÓN. VIII CONGRESO NACIONAL

TELEDETECCION, 217-219.

Modig, S.A., 1992. Swedish Forecasts of sugar beet yields -Some Regression Models. In: Proceedings of the /IRB 55th winter congress, pp. 189–210.

Nigro, Héctor Oscar (2014). KDD (Knowledge Discovery in Databases). UNICEN, 12-16.

Ofer Beeri, X. Z. (2004). Using Landsat Images To Map Quality And Quantity Sugar Beet Yield. Sugarbeet Research and Extension Reports, 125.

P. López, A. C. (2005). Metodología operativa para la obtención del coeficiente de cultivo desde imágenes de satélite. Dialnet, 212-224.

Panella, L., Kaffka, S.R., Lewellen, R.T., McGrath, J.M., Metzger, M.S., Strausbaugh, C.A., 2014. Yield Gains in Major U.S. Field Crops. CSSA Special Publication, pp. 357–395.

Qi, A., Kenter, C., Hoffmann, C., Jaggard, K.W., 2005. The Broom's Barn sugar beet growth model and its adaptation to soils with varied available water content. Eur. J. Agron. 23, 108–122.Raes, D.; Steduto, P.; Hsiao, T.C.; Fereres, E. AquaCrop—The FAO Crop Model to Simulate

Russello, H. (2018). Convolutional neural networks for crop yield prediction using satellite images. IBM Center for Advanced Studies.Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, Pages 53-65.

Ruß, G. (2009). Data Mining of Agricultural Yield Data: A Comparison of Regression Models. SpringerLink, 24-37.

Sadaf Hossein Javaheri, B. T. (2014). Data Mining Applications with R.

Spitters, C.J.T., van Keulen, H., van Kraalingen, D.W.G., 1989. A simple and universal crop growth simulator: SUCROS87. In: Rabbinge, R., Ward, S.A., van Laar, H.H. (Eds.), Simulation and Systems Management in Crop Protection. Wageningen, pp. 434 Goudriaan, J. & H.H. van Laar, 1994.

Steduto, P.; Hsiao, T.C.; Raes, D.; Fereres, E. Aquacrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. Agron. J. 2009, 101, 426–437.

Taky, A., 2008. Maîtrise des excès d'eau hivernaux et de lirrigation et de leurs conséquences sur la productivité de la betterave sucrière dans le périmètre irriguédu Gharb (Maroc). In: Analyse expérimentale et modélisation. AgroParisTech, France Ph.D. Thesis.

Vandendriessche, H.J., van Ittersum, M.K., 1995. Crop models and decision support systems for yield forecasting and management of the sugarbeet crop. Eur. J. Agron. 4 (3), 269–279.

Vandendriessche, H., 1989. Het suikerbietenmodel SUBEMO. In: Simulatie als hulpmiddel bij het stikstofbemestingsadvies voor de teelten wintertarwe en suikerbieten (I.W.O.N.L.). pp. 83–108.

Vandendriessche, H.J., 2000. A model of growth and sugar accumulation of sugar beet for potential production conditions: SUBEMOpo I. Theory and model structure. Agric. Syst. 64, 1–19.

Yield Response to Water: II. Main Algorithms and Software Description. Agron. J. 2009, 101, 438–447.

You, J., Li, X., Low, M., Lobell, D., and Ermon, S. (2017). “Deep gaussian process for crop yield prediction based on remote sensing data,” in Thirty-First AAAI Conference on Artificial Intelligence (San Francisco, CA), 4559–4566.

Webb, C.R., Werker, A.R., Gilligan, C.A., 1997. Modelling the dynamical components of sugar beet. Ann. Bot. 80, 427–436.