USE OF DESIGN OF EXPERIMENTS, DESIGN THINKING, AND DATA SCIENCE IN PRECISION AGRICULTURE

Verónica SAAVEDRA, Carlos GONZALEZ-ALMAGUER, Arturo BARRON, Claudia ZUBIETA, Eduardo CABALLERO, Hugo PEREZ and Lourdes MUCIÑO
Tecnologico de Monterrey, Mexico

ABSTRACT
Precision agriculture was born with the massive introduction of GPS at the beginning of this century and consists of the use of technology to obtain data, its statistical treatment, and decision-making.
At the Tecnologico de Monterrey experimental agriculture centre, precision agriculture has been implemented to obtain better sowing techniques and to optimise resources; this information has served the farmers of the Bajío region to improve their harvests.
The use of satellites to obtain sensitive information, in addition with technological innovation such as the use of sensors and drones, provides a large amount of information that requires the implementation of design techniques. First, it is necessary to build databases and later, through the design of experiments and design thinking, to elaborate the statistical experiments in which the appropriate parameters will be obtained to maximise the production of agricultural products with the highest nutritional value and quality.
Nowadays, supported with data science, design of experiments, and precision agriculture the present research is the base for future works applying these concepts in order to contribute to the development of Mexican agriculture. This has led during the last year that the department of industrial engineering of the School of Engineering and Sciences of Tecnológico de Monterrey the need of the design of databases to take advantage of the information obtained in the experimental field and using dynamic factor factorial design techniques and machine learning, they were capable to obtain information for decision-making.

Keywords: Design of experiments, innovative education, data science, precision agriculture, higher education

1 INTRODUCTION
This paper shows the results obtained from the design of a methodology that fuses the Design of Experiments, Design Thinking to Data Science, as well as its implementation for the subject of Analysis of Design of Experiments, and the end-of-course projects were made at the Centro Agrícola Experimental del Tecnologico de Monterrey (CAETEC) focused on the search for the appropriate parameters to maximise the level of nitrogen in corn, which is part of the diet of cattle in milk production. Likewise, CAETEC is a living laboratory focused on the practice of techniques and theory seen in class to maximise the experiential learning of students according to the TEC 21 Model of the Tecnologico de Monterrey (challenge-based learning).
In 1564, Vasco de Quiroga began technical agriculture in the new world by comparing climates and soil characteristics in Spain to introduce new crops in the Purépecha region today known as Michoacán, with which he took advantage of natural conditions to generate poles of development through agriculture. Sir Ronald Aylmer Fisher was a statistician and biologist. In 1919 he began working at Rothamsted Research, an agricultural experiment station, where he developed the analysis of variance to analyse his vast data on crops grown since the 1840s, which gave rise to the design of experiments.
The design of experiments refers to the process of designing and conducting experiments in such a way that adequate data are collected that can be analysed with statistical methods to obtain valid and objective conclusions. (Montgomery, D.) [1] A natural starting point for the design of experiments in agriculture

is the excellent review by Verdooren (Verdooren, LR) [2] who tracks development from ancient times through to the advent of Sir RA Fisher in 1926, which represents the start of modern statistical design. The design of experiments can be used as a statistical simulation tool, in which the parameters can be changed, and different results observed; in agriculture it allows us to approach the ideal parameters without having to wait for the sowing - harvest cycle to end to analyse the results, this allows us to dramatically reduce research and development time (Almaguer, C. et al) [3]. Figure 1 shows the precision agriculture timeline.

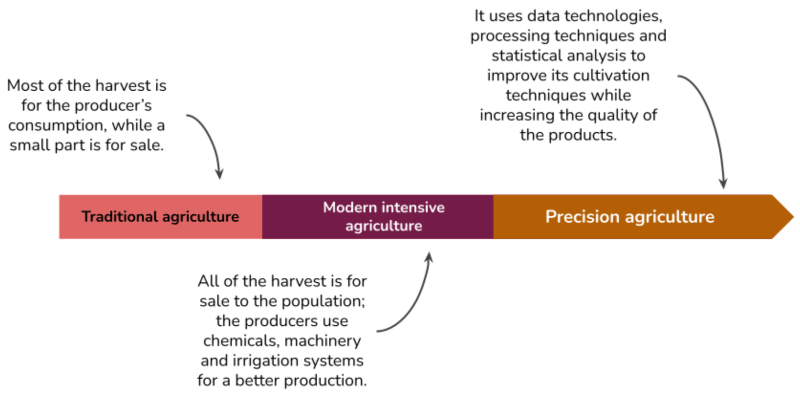

Figure 1. Timeline of precision agriculture

For students, the most difficult part in the design of experiments process is the definition of the problem, abstracting reality into a model where the variables and interactions of the system can be reflected is not very simple, to correct this weakness in 2009, González and Lloveras [4] proposed to introduce the engineering of thought that is based on the algebra of sets to focus the investigation only on the area of interest. This methodology was enriched with the introduction of design thinking and in this way, we proposed a methodology that fuses this technique with the design of experiments primarily to generate learning activities but has led to the design of new products or processes. Figure 2 shows the relationship between Design Thinking, Design of Experiments, and Data Science.

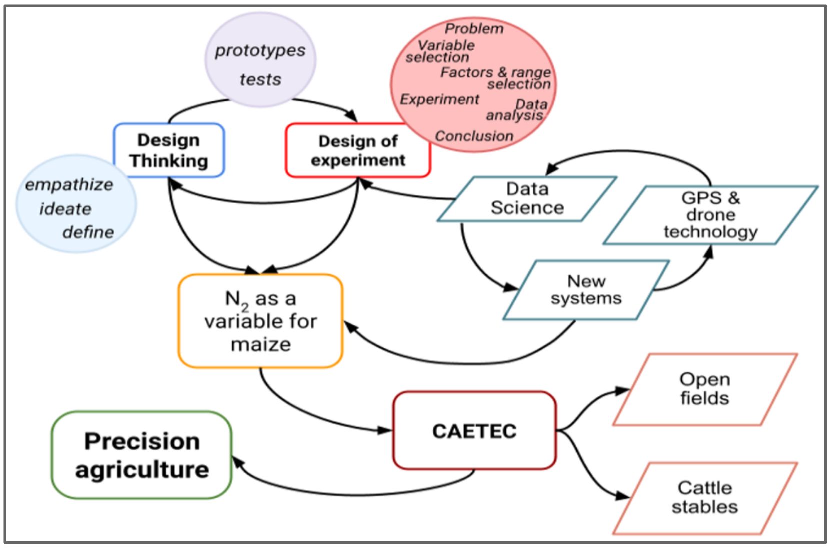

Figure 2. Relationship among design thinking, design of experiments and data science Self-elaborated

Design thinking has been used to educate in medical schools with two primary functions: the first, to develop a specific new product and the second, to develop in the students a problem-oriented way of thinking about the development of a new product. Regarding the last, Hernández [5]. Sanders J. and Goh P. [6] illustrate design thinking among second-year students to develop a new community service, highlighting as challenging the stakeholders' diverse perspectives and the introduction of extra activities

to the medical students' curriculum. Design thinking has also benefited the business administration program at the University of Amazonia, where students generate experiences corresponding to their reality as citizens. By applying DT principles, they can promote the learning of social thinking skills in other people. (González-Almaguer, C., et al)[7].
The term Data Science (DS) refers to an interdisciplinary field that involves a series of methods, processes, and systems, with the aim of extracting knowledge from data. DS, which is a discipline very related to Computing, has proved to be of great application in very different domains, particularly Education (Klašnja-Milićević et al., 2017) [8].
The application of DS in the field of Education may result in great interest for involved stakeholders (students, instructors, institutions, etc.) since the extracted knowledge from educational data would be useful to deal with educational problems such as students' performance improvement, high churning rates in educational institutions, learning delays, and so on. There are a series of disciplines related to Educational Data Science, such as Educational Data Mining and Learning Analytics (Romero & Ventura, 2020), [9], and all of them are of importance for this special issue.

2 DOE IN PRECISION AGRICULTURE

CAETEC is an educational laboratory for the students of the Tecnologico de Monterrey, and it is the place where they are expected to learn about the different types of fertilisers, the types of soil, and the type of irrigation used and put into practice the concepts acquired to the experimental design; therefore, the implementation of precision agriculture can be a success factor in practice. 'Precision agriculture is a management strategy that gathers, processes, and analyses temporal, spatial, and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production.'(International Society Precision Agriculture) [10]. For academic purposes, the study of this branch of knowledge is integrated as an enterprise resource planning system (ERP) and the generation of the data is the raw material for the design of experiments, as well as the data science. Integrated information is the source for precision agriculture, and it helps researchers and students to develop a prospective thinking.
To maximise milk production, the significant variable is the level of nitrogen in corn [11], which is an essential component in the diet of cows, being the main factor for milk production. The cycle begins with a corn seed and ends with the production of milk.
To achieve these conditions, it is necessary to experiment with different planting methods to obtain corn, which in turn will be the primary food for the cows. CAETEC has a cattle herd focused on milk production. For this reason, it is intended to carry out the design for precision agriculture with which it is sought to maximise the level of nitrogen in corn. For this analysis, a design of experiments of three factors was proposed: Fertiliser, Type of irrigation, and Type of soil. These factors are considered by CAETEC as fundamental in the percentage of nitrogen in corn. With the approach of the design of experiments, it is intended to validate this argument.
In this way, students will not only learn the different types of fertilisers, type of soil, and type of irrigation, but they will also put into practice the concepts acquired in the experimental design. The following figure shows the relationship between the three factors to be analysed and their interactions.

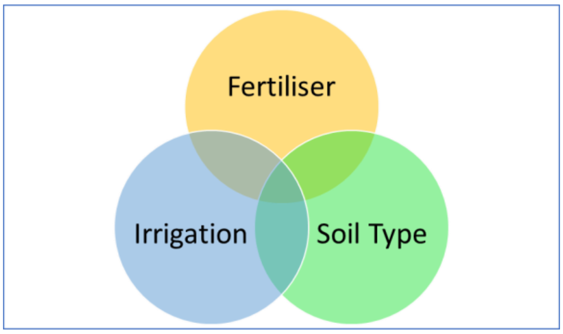

Figure 3. Example of Three factor Factorial Design. Self-elaborated.

2.1 Proposal for design of experiments

Halamchi, et al. [12] proposed Regression Analysis for predicting dry matter intake (DMI) in cows, this affects directly in the production of milk. The nature of the data in CAETEC does not allow the use of regression analysis and hence Design of experiments was proposed.

According to Montgomery [1] the guidelines for Designing and Experiment is given by:

1. Recognition of and statement of the problem Pre-experimental
2. Selection of the response variable
3. Choice of factors, levels, and ranges
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of the data
7. Conclusions and recommendations

The purpose of this article is to design jointly with CAETEC and the School of Engineering and Sciences of the Tecnologico de Monterrey until step 4, that is, until the selection of the experimental design.

2.2 Three-factor factorial design

The first proposal for CAETEC consisted of using three factor factorial design which involves three factors or sets of treatments. Factor A was going to be the soil type, Factor B was going to be the fertiliser, and Factor C was going to be the irrigation method, these are arranged in a factorial design; that is, each replicate of the experiment contains all treatment combinations. For this type of design, it is necessary to have at least two replicates of the treatments. The order in which the observations are going to be taken will be at random so that this design will be a completely randomised design.

Factor A (Soil type)

Factor B (Fertiliser)

Factor C (Irrigation method)

The effect model i was given by:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \qquad (1)$$

where:

□ is an overall mean effect,

τ_i is the effect of the ith level of the row factor A,

β_j is the effect of the jth level of column factor B,

γ_k is the effect of the jth level of the row Factor C,

$(\tau\beta)_{ij}$ is the effect of the interaction between i and j

$(\tau\gamma)_{ik}$ is the effect of the interaction between i and k

$(\beta\gamma)_{jk}$ is the effect of the interaction between j and k

$(\tau\beta\gamma)_{ijk}$ is the effect of the interaction among i, j, and k

ϵ_{ijkl} is the random error, Normal(0,□^{2})

All the factors are assumed to be fixed. The treatment effects are defined as deviations from the overall mean and the effects of the interactions are fixed and also add up to zero.

3 A NEW MODEL TO OPTIMISE THE RESPONSE VARIABLE

Traditionally we conceptualise the factors, as referred to in figure 2, and we look for their relationships, in a said model that reflects traditional agriculture we have 3 factors that, in turn, represent 7 hypotheses of interest, being the dominant one or, the one that explains the response variable is the intersection of the three, and for this, we must determine the levels until we find the perfect combination. There is also a need to analyse new factors such as the type of seed and the nutrition of the crop. In the CAETEC the students and experts, thanks to design thinking, noted that this kind of model was insufficient to predict the response variable, hence a new methodology was proposed.

Design is everything. New technologies and the exponential growth of data science challenge researchers in new ways to conceptualise factor selection and levels to maximise response variable, and the paradigm shift is to migrate from static design of experiment models to dynamic experiment models through data science, specifically with the use of machine learning algorithms that instantly generate the necessary levels to generate a design of experiments and maximise the response variable at a point in time.

The use of state-of-the-art technology to obtain information such as drones, sensors on tractors, and satellites, allows us to obtain data in real time, which, when used in correct models, allows us to make the correct combination of levels in the factors that give as a result, the optimization of the NDVI level. The NDVI is a dimensionless index that describes the difference between visible and near-infrared reflectance of vegetation cover and can be used to estimate the density of green on an area of land (Weier and Herring, 2000) [11], which in the end we can translate as the nitrogen level.

In this new model, we conceptualise all this information to obtain the levels, we must change towards a view of decision trees in a timeline, that is, each factor is, in turn, made up of multiple variables that are derived, in turn, in multiple experiments. For example, to form the favoured level, we require information from the soil, and this is formed by the depth, length, and density, the information obtained by the satellites together with the combination of the sensors in the tractors allows us to have relevant data on the mineralization of the land, combined with the time variable, and the region or cell studied (a cell studied is the coordinate obtained in a crop field) generates its levels.

The design of machine learning algorithms will allow us to generate and process this information to predict the correct factors and levels for the factorial design. This process will be replicated for all the necessary factors in order to maximise the nitrogen in the corn.

The use of statistical quality control tools, such as the Pareto diagram, will allow us to define which factors are the most decisive to be analysed and simplify their application in the agricultural field. A dynamic design of experiments will allow us to make corrections in the timeline of the sowing-harvesting process, as well as to do preventive measures when the machine learning algorithms give us warnings about external factors that affect our production system, such as the climate and pathogens, and will allow us to prevent these variables from affecting the result. Figure 4 shows us the proposed methodology.

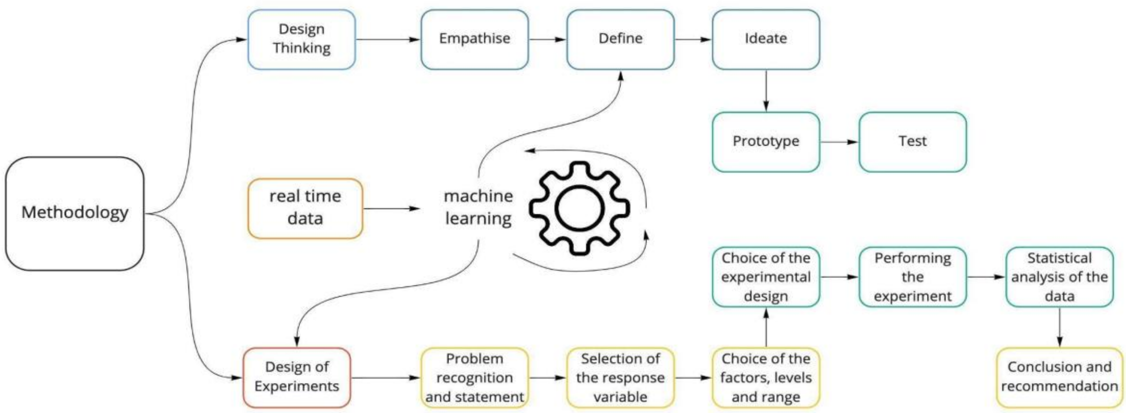

Figure 4. Proposal of Methodology. Self-elaborated.

The design of the layout for the databases is decisive, this will optimise the data mining process, facilitating the selection of quality data and thus avoid contamination of information by useless data; in this part, statistics play a fundamental role.

4 CONCLUSIONS

In the present research, we worked on the design and definition of the variables of interest to improve the corn crop, that is, up to the approach of the problem and the suggested methodology for the analysis of the information to be able to carry out the CAETEC data science. It is necessary to work with algorithms that allow us to find the levels of the factors in the proposed design of experiments. Once the data is collected, the DS will be crucial in the analysis of the information. DS added to the design of experiments will give to CAETEC a powerful tool to making decisions to maximise the nitrogen in corn to feed the cows. The Bajío Region in Mexico will also be benefited with this analysis.

Due to the sowing time and the different iterations or improvements of the parameters of the critical factors. Final results that impact until the maximisation of milk production will be presented in future research.

ACKNOWLEDGEMENT

The authors acknowledge the technical and financial support of the Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in the production of this work.

REFERENCES

[1] Montgomery D. *Diseño y análisis de experimentos*, 2011 (Limusa Wiley, México).

[2] Verdooren L. R. History of the Statistical Design of Agricultural Experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2020, 10.1007/s13253-020-00394-3.

[3] Almaguer C., Montejano G., VázquezFeijoo H. and Nieto Barrera J. Proyectos de Desarrollo Rural en Zonas Marginadas de Mexico Usando Ingenieria concurrente. *Congreso Iberoamericano de Ingeniería de Proyectos*, CIIP'03, Mar de Plata, Noviembre 2012.

[4] González C. and Lloveras J. Proyectos de Desarrollo, Diseño de Nuevos Productos e Innovacion para Agroindustrias en Zonas Marginadas de Mexico. *In International Congress on Project Management and Engineering*, AEIPRO'08, Badajoz, Julio 2009.

[5] Hernández-Gil C. and Núñez-López J. A. Design thinking aplicado al mejoramiento de las competencias ciudadanas en universitarios: voto popular. Revista de Investigación Desarrollo e Innovación, 2020, 11(1), 85-98. *https://doi.org/10.19053/20278306.v11.n1.2020.1168*.

[6] Sandars J. and Goh P. S. Design Thinking in Medical Education: The Key Features and Practical Application. *Journal of Medical Education & Curricular Development*, 2020, 7, 1-5.

[7] González-Almaguer C., Saavedra V., Caballero E., Acuña A., Zubieta C., Barbosa E. and Lule M. Design Thinking and Design Experiments: The Fusion of the School of Design and Industrial Engineering to Create Learning Experiences in The Tec21 Educational Model. In *International Conference on Engineering and Product Design Education*, EPDE 2021, Herning, September 2021.

[8] Klašnja-Milićević A., Ivanović M. and Budimac Z. Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education,* 2017, 25, 1066–1078. https://doi.org/10.1002/cae.21844

[9] Romero C. and Ventura S. Educational data mining and learning analytics: An updated survey. *Wires Data Mining and Knowledge Discovery*, 2020, 10, e1355, https://doi.org/10.1002/widm.1355

[10] International Society Precision Agriculture. Precision Ag Definition. Available: *https://www.ispag.org/about/definition* [Accessed on 2022, 04 March] (2019).

[11] ElMaraghy W., ElMaraghy H., Tomiyama T. and Monostori L. Complexity in engineering design and manufacturing. *CIRP Annals-Manufacturing Technology*. 2012. ;61(2):793-814.

[12] Halachmi I., Edan Y., Moallem U. and Maltz E. (2004). Predicting feed intake of the individual dairy cow. *Journal of Dairy Science*, *87*(7), 2254-2267.