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

Teaching Mathematics towards Precision Agriculture through Data Analysis and Models

R. F. Sviercoski

Department of Mathematics - Oklahoma University – Norman – OK

A paper for the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract.

Precision agriculture is used in a wide variety of field operations and agricultural practices that affect our daily lives. Many fields of agriculture are increasingly adopting equipment automation, robotics, and machine learning techniques. These all lead to recognize that data collection and exploitation is a valuable tool assisting in real-time farming and livestock decisions. Thus, the immediate need to empower students in Agriculture Sciences with mathematical tools using data analysis is more imperative than ever before. Such tools will better prepare them for the challenges they face while working with data analysis and models in the context of Internet of Things (IoT). This paper aims to summarize previous experience in teaching Mathematics using data analyses as motivation to important Mathematical concepts towards applications in the area. This is a result from a thorough research and discussion with colleagues among many areas of applications within the field of Agriculture Sciences, which is in itself a rich interdisciplinary field. The examples gathered from this work were then used as motivation for teaching Mathematics to agriculture engineering students based and data analysis, aimed to prepare them towards using these tools more effectively not only throughout their degree, but helping to advance the field towards precision agriculture. I will describe a few instances of how data obtained from the literature are contextualized to present the concepts of functions, derivative and integration, multivariable functions, and linear algebra. In this one-year course, the students were introduced to and got familiarized with concepts and laws that would help them into to succeed in Soil Physics, Statistics, Meteorology, Topography, Economics, Biology, and so on. Besides sharing this experience, the idea here is to stir in the community this new paradigm for teaching Mathematics in the field of Agriculture Sciences.

Keywords. *Teaching Mathematics, precision agriculture, data analysis, models.*

Introduction

Agricultural Sciences is possibly the field that most directly influences our daily lives. That is because it deals with our food supply and security. Today precision agriculture is used by a wide variety of farming operations and agricultural practices. This is primarily due to the reduction in the cost of acquisition of data and technology making precision agriculture a viable path to make agriculture more sustainable and increasing food availability.

For example, precision agriculture seeks to use new technologies to increase crop yields and profitability while lowering the levels of traditional inputs needed to grow crops (land, water, fertilizer, herbicides and insecticides). Fields can be leveled by lasers, which means water can be applied more efficiently without running off into local streams and rivers. Field precision agriculture is enabled by using data obtained by precision geolocation, remote sensing, and the mapping of soils, nutrients, and crop conditions as well as the location of weeds and pathogens.

This implies that an agriculture engineer must be prepared to make better use of this abundance of data and apply it to a more effective decision making.

This also calls for a better multidisciplinary integration between Mathematics and Agronomy aiming to equip students with critical thinking about the use of data analysis and mathematical models in their area of application. By teaching these students, since their first year of undergraduate education, Mathematics with motivation focused on data analysis and models concerning their area of application, will better prepare them for their classes that uses these concepts and will encourage them to develop more scientific questions that turn into future investigations and development of new research areas in their filed.

This type of approach is nowadays more common in Mathematics as well, that is why the idea of the STEM (Science, Technology, Engineering and Mathematics) education came about. However, the difficulty with the use of mathematics for agricultural scientists is that most courses in applied mathematics have been designed for other engineering students and without the data analysis. Agronomy, however, is different than the other engineering fields, because it comprises the use and application of biology, chemistry, plant genetics and physiology, animal sciences, meteorology, topography, soil sciences, economy and administration. Meaning that it is itself a multidisciplinary field, thus the need to propose an approach that will make use of this characteristic.

What will be described are some examples on how one can present and explore mathematical concepts using data and the idea of creating a mathematical model for prediction. This idea was first developed in a research proposal in Brazil that produced a textbook, on its 8^{th} reprint now, which can be translated as: Applied Mathematics to Agriculture Sciences: Data Analysis and Models [2]. This work is now being reviewed with the intend of translating to English and extending to include further advanced topics like Differential Equations.

Calculus (in one variable) using Data Analysis

Using data sets to motivate mathematical models is based on assumptions that each data set presented describes an exact mathematical relationship called function. By analyzing the phenomenon one can establish the domain and range of such functions as well whether this function (phenomenon) describes an increasing or decreasing behavior. It will be presented an example and how this can be dealt in the context of Agronomy.

Basics Functional Models:

As an example, a polynomial model will be presented here, because it comprises concepts used in derivatives and integral. There are similar techniques for exploring data representing exponential, logarithm, power function and trigonometric models.

In order to identify whether a data set describes **polynomial behavior**, one can compute and analyze whether the first rate of change (FRC). This information is crucial in data analysis because it determines how fast (or slow) the function is increasing or decreasing. If FRC is constant, then the data describe a linear function. If not, one can proceed to compute the second rate of change (SRC), to verify maybe whether this data models a quadratic function, or yet use the third rate of change (TRC) to verify whether it is a cubic, and so on. Obviously, it is very important to understand what each rate of change tell us about the data, and a sketch of its graph. A data set modeling a cubic polynomial summarizes this information on Example 1.

Example 1: The first 2 columns of Table 1 describe corn production *C* [tons/acre] as a function of phosphorus *p* [tons/acre] added as fertilizer. The FRC, SRC and TRC are obtained as:

Table 1. Corn production *C [t/a]* as a function of amount of phosphorus *p [t/a]* –

p[t/a]	*C(p) [t/a]*	***FRC*** **=** $\frac{\Delta C}{\Delta p}$	**SRC** = $\frac{\Delta^2 C}{\Delta p^2}$ $[t/a]^{-1}$	**TRC=** $\frac{\Delta^3 C}{\Delta p^3}[t/a]^{-2}$
0	72.15			
2	88.15	8		
4	109.75	10.8	1.4	
6	134.55	12.4	0.8	-0.3
8	160.15	12.8	0.2	-0.3
10	184.15	12	-0.4	-0.3
12	204.15	10	-1.0	-0.3
14	217.75	6.8	-1.6	-0.3
16	222.55	2.4	-2.2	-0.3
18	216.15	-3.2	-2.8	-0.3

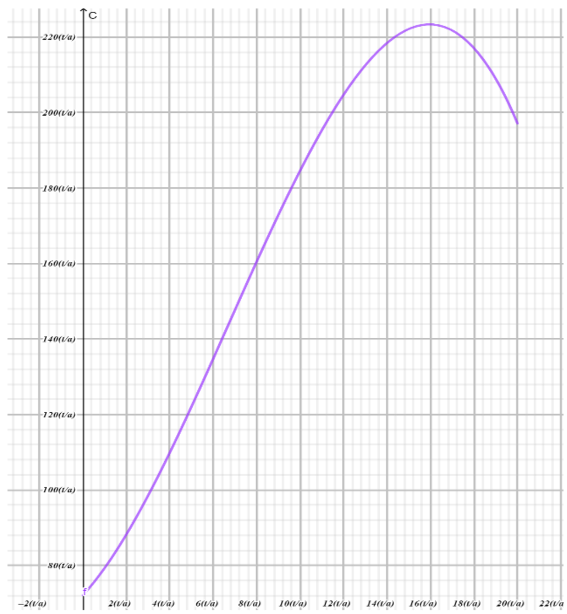

Figure 1: Graph obtained from Eq. (1)

After this analysis and by working backward substitution as a simple system, one can conclude that the corn production given by this data is described by

$$C(p) = -0.05p^3 + p^2 + 6.25p + 72.15 \; [t/a] \quad (1)$$

Other similar mathematical strategies are available to identify whether a data can be modeled by **polynomial of other degrees**, **exponential, power functions or trigonometric functions**. **Limit** (including left and right) and asymptotic behaviors are introduced from data describing **rational functions, exponential with limited growth**, and **logistic function**.

The crucial learning objective is to be able to identify which functional behavior each data set describes, the meaning of the FRC and what further info it gives about the data, similarly for the SRC, the TRC and so on. Also, based on these analyses, one should be able to sketch a graph.

2.2 Derivatives of various orders

The data analysis previous described sets one up to introduce the definition of **derivative** which comes naturally from the FRC. However, in this context, the motivation stirs from the need of a tool that would estimate values that are not given on the table. For example, from the third column of Table 1, one can conclude that for phosphorus addition between $p = 6\ [t/a]$ and $p = 10\ [t/a]$, the production responses are the highest (fastest). Because one wants an optimal response in production, it would be fair to predict the production at $p = 7\ [t/a]$ or at $p = 9\ [t/a]$.

After this analysis, one should recognize the limitations of the approximation and so the need to introduce the definition of derivative, using the formalism of Limit and further apply to the analytical Equation describing the data. One can return on each of the examples behind and motivate the need for concise rules for derivatives. Note that Derivative is the direct application of what is called **variable rate technology** (VTR) [2], built into farm machinery, that yields monitoring, auto-steering, navigation using on-board computers and network capability.

The second derivative also comes as a natural consequence of analyzing further the data to convey the concept of concavity, maximum and minimum, as well as the definition of inflection points. These concepts are of extreme importance for applications of mathematics in many fields. Let's keep in mind, again, that agricultural science is a multidisciplinary field, and such concepts have well known applications in economy, biology and chemistry to name a few.

2.3 Integration

The concept of integration will be illustrated in Example 2 where the soil moisture $\theta\ [cm^3/cm^3]$ depends on the depth of soil profile $z[cm]$.

Observe also that the students get familiarized with the formulation and techniques that will help them into Soil Physics and Meteorology.

Example 2: Table 2 describes soil moisture $\theta\ [cm^3/cm^3]$ depending on depth $z[cm]$.

Table 2: Soil moisture θ as a function of depth z

$z[cm]$	$\theta(z)\ [cm^3/cm^3]$	$FRC = \frac{\Delta\theta}{\Delta z}\ [cm^{-1}]$
0	0.15375	0.0012
15	0.17160	$5.200E^{-4}$
30	0.17940	$-3.0666E^{-4}$
45	0.17480	0.0020
60	0.20460	-0.00169
75	0.17930	-0.00236
90	0.14385	

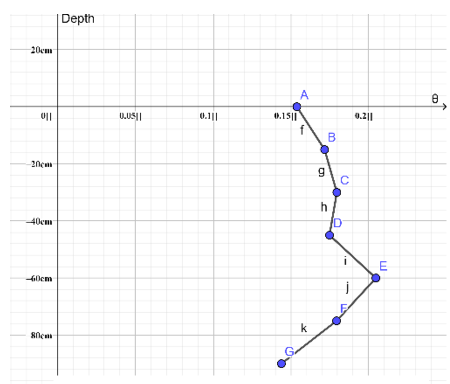

Figure 2: Soil moisture θ as a function of depth (z)

In order to compute the amount of soil moisture accumulated, or water stored W_s, from 0-90 cm, we use the integral:

$$W_s = \int_0^{90} \theta(z)\, dz \tag{2}$$

This is a classic example where one can introduce the concept of integration. At first, again one can start with the numerical integration, then moving into the analytical.

2.3.1 Density Function and Cumulative Density Function

Another important application of integration has to do with understanding how a characteristic of a population is distributed within the population, one resorts into understanding the distribution and density of such characteristic, which will then use integration to compute probability.

Example 3: Consider a plantation's plot of corn, and one wants to understand how the height of each plant is distributed inside the plot, which is designed in Figure 3,

206	230	229	247	235	260	240	233
217	236	243	195	249	232	244	257
233	235	241	225	217	230	252	239
232	268	239	208	240	233	210	258
190	232	219	245	196	216	232	235
239	212	229	201	233	254	225	249
222	244	228	221	232	215	227	198
216	247	250	200	239	243	237	225
228	226	253	225	247	244	233	203
213	222	240	268	226	221	215	232

Figure 3. Experimental plot of Corn plantation

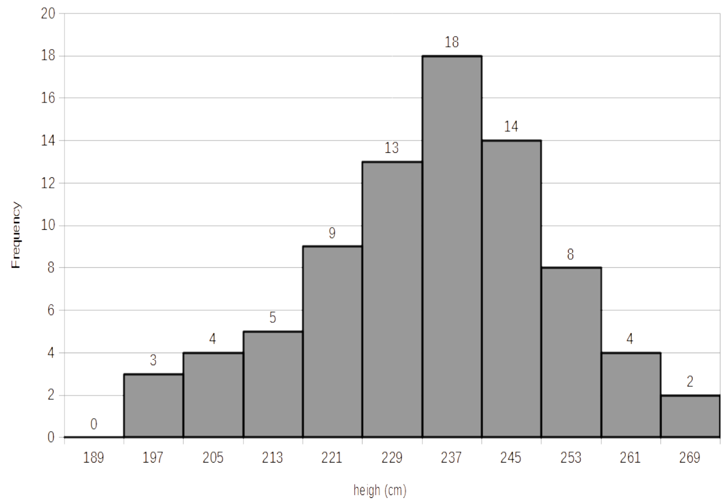

Table 3: Height's distribution from Fig.3

h =Intervals (cm)	Frequency	Density (%)	$D(h)$ =Density $\frac{\%}{cm}$
189	0	0	0
197	3	3.75	0.46875
205	4	5	0.62500
213	5	6.25	0.78500
221	13	11.25	1.40625
229	18	16.25	2.03125
237	14	22.5	2.8125
245	12	17.5	2.18750
253	8	10	1.25000
261	4	5	0.62500
269	2	2.5	0.25000

Figure4:Distribution of the height of corn plot Fig. 3

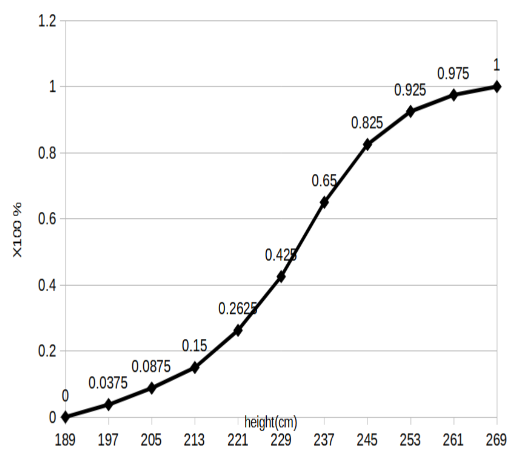

Figure 5: Cumulative Distribution

From Table 3, the histogram (Fig.4) is obtained, from which one can also obtain a **probability function**, which can help us calculate the probability of obtaining from the crop, a plant with height between $h = a$ and $h = b$ as:

$$P(205 \leq h \leq 229) = \int_{205}^{229} D(h)dh = 5 + 6.25 + 11.25 + 16.25 = 35.75\% \quad (3)$$

Which can be computed by numerical integration methods like Riemann Sums, Midpoint or Trapezoidal method, or another. All those methods are useful when using data, because each has properties that are suitable to one or other type of data distribution.

Another direct application of integration is the **cumulative distribution function** (CDF), illustrated in Fig. 5, which is defined as

CDF=$P(y)$ is the percentage of plants whose height is less than $y = \int_0^y p(x)dx$ (4)

Calculus of Two or More Variables

To introduce functions of two or more variables, again it can be done by using data from experiments, then the definitions of domain, range need to be revised in this context. The derivative and maximum and minimum also take new notations, but all based on the knowledge and a generalization of the one variable. By this time, it should already be clear the meaning and what information we gather by making these computations.

Example 4: Table 4 registers *P(x,y)* as the production of beans (*Phaseolus vulgaris L.) (kg/ha)* by changing levels of Nitrogen *(kg/ha),* which we set to be our variable *x*, and water level *(mm),* which is our variable *y*. The **Domain** can be described by the Cartesian interval notation $D = [0,220]X[105,635]$. From the data, one obtains the 3-D graph, as Fig. 6:

Table 4: Production of Beans as function of x= Nitrogen (kg/ha) and y= water (mm)

$y\backslash x$	0	20	40	60	80	100	120	140	160	180	200	220
105	1500.937	1751.757	1929.537	2034.277	2065.977	2024.637	1910.257	1722.837	1462.377	1128.877	722.3365	242.7565
145	1733.937	1996.917	2186.857	2303.757	2347.617	2318.437	2216.217	2040.957	1792.657	1471.317	1076.937	609.5165
180	1915.394	2189.014	2389.594	2517.134	2571.634	2553.094	2461.514	2296.894	2059.234	1748.534	1364.794	908.014
225	2117.953	2405.253	2619.513	2760.733	2828.913	2824.053	2746.153	2595.213	2371.233	2074.213	1704.153	1261.053
265	2268.969	2568.429	2794.849	2948.229	3028.569	3035.869	2970.129	2831.349	2619.529	2334.669	1976.769	1545.829
305	2392.657	2704.277	2942.857	3108.397	3200.897	3220.357	3166.777	3040.157	2840.497	2567.797	2222.057	1803.277
345	2489.017	2812.797	3063.537	3241.237	3345.897	3377.517	3336.097	3221.637	3034.137	2773.597	2440.017	2033.397
385	2558.049	2893.989	3156.889	3346.749	3463.569	3507.349	3478.089	3375.789	3200.449	2952.069	2630.649	2236.189
425	2599.753	2947.853	3222.913	3424.933	3553.913	3609.853	3592.753	3502.613	3339.433	3103.213	2793.953	2411.653
465	2614.129	2974.389	3261.609	3475.789	3616.929	3685.029	3680.089	3602.109	3451.089	3227.029	2929.929	2559.789
505	2601.177	2973.597	3272.977	3499.317	3652.617	3732.877	3740.097	3674.277	3535.417	3323.517	3038.577	2680.597
550	2553.94	2940.04	3253.1	3493.12	3660.1	3754.04	3774.94	3722.8	3597.62	3399.4	3128.14	2783.84
595	2472.117	2871.897	3198.637	3452.337	3632.997	3740.617	3775.197	3736.737	3625.237	3440.697	3183.117	2852.497
635	2370.349	2782.289	3121.189	3387.049	3579.869	3699.649	3746.389	3720.089	3620.749	3448.369	3202.949	2884.489

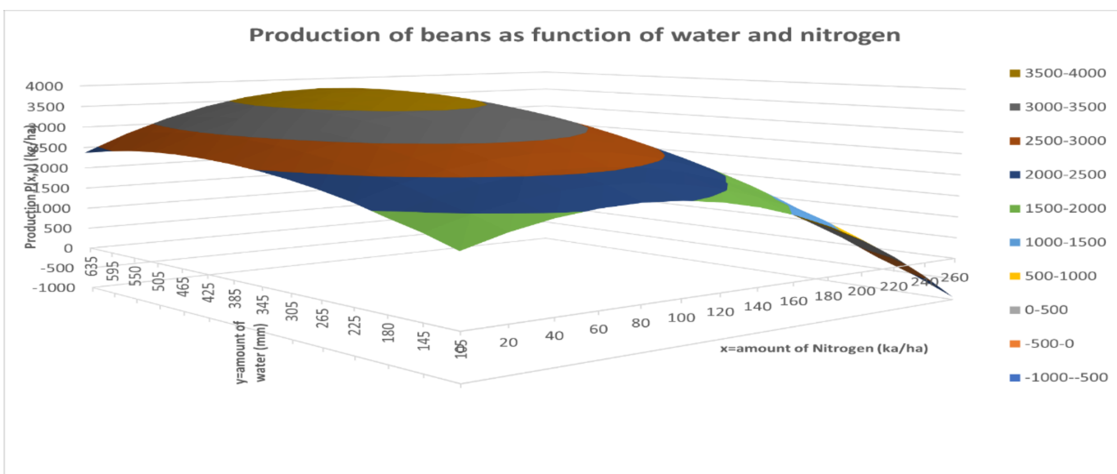

Figure 6: Graph obtained from the data on Table 4.

The first new idea here is the level curves, corresponding to the combination of the variables x

and y that will yield each fixed production level. The concept of **partial derivative** can again be introduced as the generalization of FRC, as **Partial First Rate of Change** (PFRC), with respect to each variable *(x,y).* Its computation produces now two tables one for the approximate derivative with respect to x, named here (PFRCx), and another with the approximate derivative with respect to y, as (PFRCy). They are defined as:

$$PFRCx = \frac{P(x_{i+1},:)-P(x_i,:)}{x_{i+1}-x_i} = \frac{\Delta P}{\Delta x} \quad (5) \qquad PFRCy = \frac{P(:,y_{i+1})-P(:,y_i)}{y_{i+1}-y_i} = \frac{\Delta P}{\Delta y}$$
(6)

One should realize that, in this example, $1 \le i \le 11$ for *x* and $1 \le i \le 13$ for *y.*

Table 5: Snapshot of PFRCx $(kg/ha \cdot mm)$ for $0 \le x \le 120$ and $105 \le y \le 595$, using (5) from Table 4.

PFRCx	0	20	40	60	80	100	120
105	12.541	8.889	5.237	1.585	-2.067	-5.719	-9.371
145	13.149	9.497	5.845	2.193	-1.459	-5.111	-8.763
180	13.681	10.029	6.377	2.725	-0.927	-4.579	-8.231
225	14.365	10.713	7.061	3.409	-0.243	-3.895	-7.547
265	14.973	11.321	7.669	4.017	0.365	-3.287	-6.939
305	15.581	11.929	8.277	4.625	0.973	-2.679	-6.331
345	16.189	12.537	8.885	5.233	1.581	-2.071	-5.723
385	16.797	13.145	9.493	5.841	2.189	-1.463	-5.115
425	17.405	13.753	10.101	6.449	2.797	-0.855	-4.507
465	18.013	14.361	10.709	7.057	3.405	-0.247	-3.899
505	18.621	14.969	11.317	7.665	4.013	0.361	-3.291
550	19.305	15.653	12.001	8.349	4.697	1.045	-2.607
595	19.989	16.337	12.685	9.033	5.381	1.729	-1.923

Observe that the increasing and decreasing behaviors for the **Nitrogen** *x* at each given level of **water** *y*. One may think the concept PFRCx, as if *y* is kept fix. In here, it is done at each row.

By analogy the PFRCy is obtained, again just part of the whole table, is illustrated on Table 6.

Table 6: Snapshot of PFRCy [ ] for $0 \le x \le 120$ and $105 \le y \le 595$, using (6), from Table 4.

PFRCy	0	20	40	60	80	100	120
105	5.8250	6.1290	6.4330	6.7370	7.0410	7.3450	7.6490
145	4.5364	4.8024	5.0684	5.3344	5.6004	5.8664	6.1324
180	5.0639	5.4059	5.7479	6.0899	6.4319	6.7739	7.1159
225	3.7754	4.0794	4.3834	4.6874	4.9914	5.2954	5.5994
265	3.0922	3.3962	3.7002	4.0042	4.3082	4.6122	4.9162
305	2.409	2.713	3.017	3.321	3.625	3.929	4.233
345	1.7258	2.0298	2.3338	2.6378	2.9418	3.2458	3.5498
385	1.0426	1.3466	1.6506	1.9546	2.2586	2.5626	2.8666
425	0.3594	0.6634	0.9674	1.2714	1.5754	1.8794	2.1834
465	-0.3238	-0.0198	0.2842	0.5882	0.8922	1.1962	1.5002
505	-1.1809	-0.8389	-0.4969	-0.1549	0.1871	0.5291	0.8710
550	-2.0455	-1.7036	-1.3616	-1.0196	-0.6776	-0.3356	0.0064
595	-2.5442	-2.2402	-1.9362	-1.6322	-1.3282	-1.0242	-0.7202

The PFRCy is obtained by computing the FRC over the column for each fixed x.

As in 1-D case, our goal is to observe how these rates are changing.

Using this numerical approach, it is reasonable to explain that at any given point (x, y) these variations now are denoted by a vector, that approximate the **Gradient Vector** as:

$$\Delta P(x,y) = \left(\frac{\Delta P}{\Delta x}, \frac{\Delta P}{\Delta y}\right) \quad (7)$$

Which will point (highlight) to regions of increasing and decreasing of the function production, and our main interest is to identify where the Numerical Gradient approaches zero, because we want to obtain the optimal combination of Nitrogen and water that would, in this case, **maximize** the production, meaning solving the system:

$$\begin{cases} \frac{\Delta P}{\Delta x} = 0 \\ \frac{\Delta P}{\Delta y} = 0 \end{cases} \quad (8)$$

By analyzing the values one can estimate the maximum of the production to be in the subregion of the domain R=[100,120]x[505,550].

Here also can be introduced the concept of **Differential** and **Directional Derivatives** to determine the **rate of change,** as it was worked in 1-D version. The data analysis and graphs allow to work these concepts either by numerical means or later, when the correspondent analytical versions be defined. These will set also the need to determine how the function is changing in directions that are not necessarily the x-y directions, but a combination of those. Even though in this paper there is not intro to vectors and operations, like the usual ones and the inner product, these are part of prerequisite for multivariable calculus. In [2], these are worked prior to functions in multiple variables and with many interesting applications including in Topography.

We can graph both tables of the PFRCx and PFRCy to analyze the behavior of each. In this case, it can be seen that each PFRC is a plane, meaning that the original data can be fit into a quadratic polynomial and by various methods, either solving system, or doing best fit methods, one can obtain the analytical equation to be:

$$P(x, y) = 759.29 + 12.771x + 7.96y + 0.0152xy - 0.0913x^2 - 0.00854y^2 \quad (9)$$

With domain $D = [0,220]\ X[105,635]\ (kg/ha, mm)$.

Another way to check that the data is indeed from a quadratic polynomial, is to compute the approximate Second Derivative, observing that one will obtain now 4 tables (instead of 2 for the gradient) corresponding to the variation of the PFRC on each coordinate *x* and *y*, meaning that one would have four **Partial Second Rate of Change** (PSRC), that would be PSRCxx, PSRCxy, PSRCyx, PSRCyy. Here one can identify that for each point, the PSRC is a 2x2 matrix:

$$\Delta^2 P(x, y) = \begin{pmatrix} \text{PSRCxx} & \text{PSRCxy} \\ \text{PSRCyx} & \text{PSRCyy} \end{pmatrix} = \begin{pmatrix} \frac{\Delta^2 P}{\Delta x^2} & \frac{\Delta^2 P}{\Delta x \Delta y} \\ \frac{\Delta^2 P}{\Delta y \Delta x} & \frac{\Delta^2 P}{\Delta y^2} \end{pmatrix}$$

(10)

Since an analytical equation that models the data was found Eq. (9), then analytical and usual approaches are used, that can always be looked back at the results of the approximated expressions in order to motivate the definitions of the continuous approach.

Besides introducing the concept of function with more than one variable, understanding its notation and setting, one can work on **optimization,** for finding **maximum, minimum and saddle points**, and more importantly its meaning and usefulness for applications in many areas. One may verify that the maximum production $P = 3779.9\ kg/ha$ occurs at *(117.43 kg/ha, 570.55 mm)*. An application of these concepts is the **Least Square Method**, which will highlight the difference between statistical and mathematical models, and it will be used in their Statistics classes.

Constrained Optimization – Lagrange Multipliers

Another topic that is not usually covered in a standard math course is constrained optimization. However, this is another crucial topic that needs to be taught because it is the basis for many applications, including the Method Simplex that will be discussed ahead into the Linear Algebra setting. One application will be illustrated that follows up as a continuation of the previous example, once the analytical function has been founded in Eq. (9).

Example 5: Considering the data from Table 4 and its model given by Eq. (9), one may want to **maximize production** subject to the fixed cost of Nitrogen and water given by the equation $C(x, y) = 1.5x + y = \$500$.

Without going into details of how to obtain such values and observing that the corresponding **Lagrangean function** $L(x, y, \mu)$ which now has three variables, which in fact here μ is a parameter, called Lagrangean multiplier. It can be verified that the new values that satisfy the constrain of cost and maximize the production is $x = 79.665\, kg/ha,\ y = 380.50\, mm\ and\ \mu = -2.672\, kg/ha \cdot \$$. The new maximum of the production is now $P = 3450.36\, kg/ha,$ which is smaller than the previous value for unconstrained optimization, as expected. The most important result here is the interpretation of the parameter μ. In this case, by using the differential of P, one can estimate that if instead of \$500, the owner is willing spend \$501, then P will increase by $\mu = 2.672$, giving $P = 3453.032\, kg/ha$.

The generalization for considering other type of constrains follows, where each constrain is associated to a new multiplier, say for example σ. Then once the new Lagrangean is obtained with now four "variables", from which one can compute the respective weight that each constrain would carry over the production. Meaning that if the absolute value of σ is greater than μ, that would imply that the constrain associated with σ has greater impact into the production than the cost which has μ as its **rate of change** parameter.

Needles to observe that one can apply this Method by posing the problem in a reverse way, that is, how to **minimize the Cost,** $C(x, y)$, subject to a set production level. Understanding how to set up such problems and interpreting its solutions may give one the appreciation for having the analytical equation handy.

Linear Algebra and its Applications

The concepts of vector and matrices and their algebraic operations, like addition, subtraction, and scalar multiplication, come naturally from manipulating Tables with numerical data. In this section, it will describe important applications of linear algebra beyond the solution of linear systems, which is the obvious one, such as Markov chains and the Simplex Method.

Solution of linear systems

This is the topic that has the most obvious applications pertaining to the various areas of Agronomy. The Examples seen before were the set up and the solution of the gradient system leading to critical points Eq. (9) and the solution of the system to determine the equation that models a data set (such as in **Example 1**). Here is a good place to understand underdetermined or overdetermined systems, as well as systems that do not have solutions.

Markov Chains

The concept will be illustrated and discussed in an example about market distribution between companies that offers one good to the market. Usually, these setting corresponds to how to separate a limited supply. This same idea comes into play into pray-predator scheme, in genetic and in economics, to name a few.

Example 6: Consider three companies supplying milk to a market with brands X, Y, Z. The amount that each company sells varies with time, based on parameters like marketing, prices and other conveniences. Assuming that there are no other brands and that constant fractions of the consumers prefer one brand over another within the given time. This is to say that on Jan. 1st, the brands X, Y and Z have fractions x_0, y_0 and z_0 of the market, respectively. Whereas in Feb. 1st they have fractions x_1, y_1 and z_1. The question is how the market evolves in time and whether this market can be considered stable or unstable?

Solution: Consider the number of consumers *n*, and the notation a_{ii} = fraction of consumers that used the same brand over time, and $a_{i\,j}$=fraction that changed from brand *j* to brand *i* with $i \neq j$.

In the present example, $1 \leq i \leq 3$ and $1 \leq j \leq 3$.

The first assumption leads to the system for the market in Jan. 1st and Feb. 1st:

$$\begin{cases} x_0 + y_0 + z_0 = 1 \\ x_1 + y_1 + z_1 = 1 \end{cases}$$

From the second assumption, the number of consumers that X has in Feb. 1st is equal to the numbers that did not change brand plus the ones that moved from Y and Z. This can be formalized as $x_1 n = a_{1,1}(x_0 n) + a_{1,2}(y_0 n) + a_{1,3}(z_0 n)$, meaning that in Feb. 1st one would have:

$$\begin{cases} x_1 = a_{11}x_0 + a_{12}y_0 + a_{13}z_0 \\ x_2 = a_{21}x_0 + a_{22}y_0 + a_{23}z_0 \\ x_3 = a_{31}x_0 + a_{32}y_0 + a_{33}z_0 \end{cases} \quad => \quad X_1 = AX_0 \qquad (11)$$

Where the matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ and $X_0 = (x_0, y_0, z_0)^t$, where "t" stands for transpose. The matrix A has a special name called **Transition Matrix** because all its entries are such that $0 \leq a_{ij} \leq 1$ and as each a_{ij} is a fraction of a population, the elements of each column must add to 1. If in Jan. 1st the initial market is $X_0 = (0.2, 0.3, 0.5)^t$ and $A = \begin{pmatrix} 0.8 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}$ then by Feb. 1st, the distribution of the market will be $X_1 = (0.27, 0.38, 0.35)^t$ whereas in March 1st, the distribution will be $X_2 = AX_1$, which give $X_2 = (0.327, 0.398, 0.275)^t$ meaning that the market is distributed with 32.7% with brand X, 38.9% with brand Y and 27.5% using brand Z.

The equilibrium of the system is reached when applying the same operation consistently over time, and the distribution $X = (x, y, z)^t$ does not change. This means that exist a vector X such that $AX = X$, this equation leads to the system $(A - I)X = 0$ where I is the 3x3 identity matrix, with the additional constraint that $x + y + z = 1$, which can be set as:

$$\begin{cases} -0.2x + 0.2y + 0.1z = 0 \\ 0.1x - 0.3y + 0.3z = 0 \\ 0.1x + 0.1y - 0.4x = 0 \\ x + y + z = 1 \end{cases}$$

Which leads to the solution $X = (0.45, 0.35, 0.2)^t$. meaning that there will be no transition from one brand to another when brands X, Y and Z have 45%, 35% and 20% of the market, respectively.

The procedure carried out is called Markov chains. As mentioned before, this set up has many applications besides economy, like in population dynamics, genetics, meteorology, to name a few.

Linear Programming and the Simplex Method

The linear programming deals with **constrained optimization** that follows in the path of Lagrange multipliers, but **here all functions are linear**, meaning its 3D surfaces are planes if you want to set in the 3D. Recall that there in Example 5, that one had a quadratic polynomial (production) in 3D subject to a linear cost function.

The Simplex Method is possibly one of the most useful and powerful tool that Mathematicians have produced in this last century! With applications in so many areas, that Precision Agriculture has also benefited from! Its usefulness comes by the fact that it is quite simple to understand – because every function is linear - and easy to set it up in a numerical algorithm to be implemented into computers. It is a direct application of a data set, as it will be illustrated next in a rather elementary way, because the idea here is to exemplify applications pertinent.

Example 7: A poultry production of chicken need to combine calories and protein for a balanced diet for the chickens. The optimal amount consists of 3000 (Cal) from calories and minimum of 17.16% from protein. Considering that the farmer has only corn and soybean meal and each Kg of corn has 8.51% of protein and 3146 (Cal) of energy, whereas the soybean has 45.6% from protein and 2283 (Cal) per Kg, however it is possible to include 0.2 Kg of soybean on each portion of the food. Considering that the price of corn is $ 0.80 per Kg and the soybean costs $3.80 per Kg, how much of each component must be mixed to make a portion that has the minimum cost?

Solution: The Table summarizes some of the data described above.

Table 7: Data from Example 7

	Corn (kg)	Soybean Meal (Kg)	Minimum requirement
Protein (%)	8.51	45.6	17.16
Calories (Cal)	3146	2283	3000
Cost ($)	0.8	3.8	

The set up corresponds **minimize the cost function** $C(x,y) = 0.8x + 3.8\,y$ where x is the amount of corn and y is the amount of soybean, considering the minimum requirement which is translated to a system of inequalities that need to be satisfied. Thus, the problem becomes to

Minimize $C(x,y) = 0.8x + 3.8\,y$

Subject to $\begin{cases} 0.0851x + 0.456y \geq 0.1716 \\ 3146x + 2283y \geq 3000 \\ y \leq 0.2 \\ x \geq 0 \\ y \geq 0 \end{cases}$

Because of its linear nature, the inequalities determine a polygonal region in 2-D with corners (vertex), and the huge result here is that the minimum of $C(x,y)$ is attained at one of these vertex. Likewise, if one looks to maximize a function, instead of minimizing. Without detailing the fundamentals of it but understanding the Lagrange multipliers method is the key to determine the vertex that will minimize the cost and at the same time fulfill the given constraints from Table 7. By simple inspection one finds that the combination of $x = 0.994\,Kg$ and $y = 0.2\,Kg$ will have the cost of 1.51$/Kg.

Obviously, one can add more restrictions combining these 2 components, and the set up can be extended to many variables.

Conclusion:

It was presented here an approach on how to introduce data analysis and modeling for teaching Mathematics to Agriculture Engineering students. This was a summary which is part of a revision of a previous work by the author. The future work intends to include other topics of relevance as well as more advanced topics, such as Differential Equations.

Reference:

[1] Boghossian, A. et. al, ***Report: Threats to Precision Agriculture***, 2018 Public - Private Exchanges Analytic Program - Threats to Precision Agriculture (dhs.gov).

[2] Sviercoski, R. F. ***Matemática Aplicada às Ciências Agrárias: Análise de Dados e Modelos***, 7^{th} reprint, Universidade Federal de Viçosa (2008).