Geoderma 334 (2019) 90–98

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes

Guilherme M. Sanches[a,b,*], Paulo S. Graziano Magalhães[a], Henrique C. Junqueira Franco[b]

[a] University of Campinas – UNICAMP – Ph.D. Program in Bioenergy – Interdisciplinary Center for Energy Planning, 13083-896 Campinas, São Paulo, Brazil
[b] CROPMAN – Agricultural Inovation, 13080-650 Campinas, São Paulo, Brazil

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Precision farming
Data mining
Random forest
Yield monitor
Saccharum spp.

ABSTRACT

The adoption of information technology (IT) and precision agriculture (PA) has converted agricultural fields into data sources. However, the transformation of data into knowledge for decision making remains a major challenge. In the Brazilian sugarcane industry, the current use of PA technology is very far from its full potential for site-specific management, mainly because yields are not temporally or spatially monitored. The objective of the present study was to investigate the relationship between the physical and chemical properties of soils and sugarcane yield, thereby identifying the soil parameters that determine the final productivity of the field. Two sugarcane fields were monitored from 2011 to 2014. During the crop season, soil samples and yield data were collected annually. A random forest algorithm was applied to investigate the influence of different soil attributes on yield using data that were collected spatially over the study period. The results showed that the amount of available soil organic matter (OM), clay content and cation exchange capacity (CEC) are important factors impacting sugarcane yield variation. Furthermore, it was found that the temporal variability in yield is caused mainly by the variability in soil pH over the study period. The results indicated that when OM increased over time, there was greater phosphorus availability. Large volumes of spatial and temporal data, together with data mining techniques, allowed the extraction of knowledge and the creation of specific management zones in the field to support the decision-making process for producers.

1. Introduction

According to the goals set by the Brazilian government at COP21, ethanol production in 2030 is expected to be 54 billion of litres, almost double the current production levels. Sugar production will increase from 38.7 million tons to 46.4 million tons. To achieve these ethanol and sugar production goals, it will be necessary to produce 942 million tons of sugarcane per season in 2030 (CNI, 2017). These demands will promote an irreversible change in the current Brazilian sugar and ethanol framework. The sugarcane industry has great potential to replace part of the expected imports of fossil fuel with ethanol and to meet the established greenhouse gas (GHG) reduction targets. Thus, increasing the agricultural yield of sugarcane provides a more economically and environmentally sustainable alternative as producing more yield in the same planted area reduces production costs and avoids the need for new fields expansions. The current Brazilian average sugarcane yield is 72.6 Mg ha^{-1} (CONAB, 2017), far from the genetic potential of the crop, which is 300 Mg ha^{-1} (Waclawovsky et al., 2010). Achieving this yield threshold seems to be a distant possibility, but investments in technology and research can contribute significantly to reaching this goal.

The adoption of precision agriculture (PA) technologies represents a promising approach to increasing agricultural yields and reducing production costs. PA comprises several techniques and technologies for managing the spatial and temporal variability of crops, and these approaches seek to improve the yield, profitability and environmental management of fields. These benefits are essentially obtained through site-specific management that considers the spatial and temporal variability of fields. The main technologies available to PA users are yield monitors, remote and proximal soil and plant sensors associated with Global Navigation Satellite System (GNSS) positioning and geographic information systems (GIS). Among the fundamental PA tools are yield monitors, which can spatially map yields and identify problems in the fields. Although widely developed and used in grain crops (Silva et al., 2011; Arslan and Colvin, 2002), yield monitors are still rarely used in the Brazilian sugarcane industry. Examples of yield

* Corresponding author at: University of Campinas – UNICAMP – Ph.D. Program in Bioenergy – Interdisciplinary Center for Energy Planning, 13083-896 Campinas, São Paulo, Brazil.
E-mail address: guilherme.sanches@cropman.com.br (G.M. Sanches).

https://doi.org/10.1016/j.geoderma.2018.07.051
Received 23 February 2018; Received in revised form 8 June 2018; Accepted 28 July 2018
0016-7061/

monitor applications come mainly from the academy, with the first works having been produced by Magalhães and Cerri (2007).

One of the technological bottlenecks preventing PA advances in the Brazilian sugarcane industry is the lack of applicable knowledge to help farmers make the right decisions. Moreover, the literature offers fewer long-term economic and environmental studies on the adoption of site-specific management of sugarcane fields compared to those that evaluate grains (Yost et al., 2016) and citrus (Colaço and Molin, 2017). The development of appropriate decision support systems for decision making remains a major hindrance to the full adoption of PA (McBratney et al., 2006). At the strategic and tactical levels, the data gathered on the performance of various farm management systems should be grouped by time to build a systematic database, allowing for "quick and preliminary" assessments of the effects of management strategies based on experiences obtained elsewhere in similar soil conditions (Bouma et al., 1999). To overcome this challenge, information technology (IT) has been widely applied in all aspects of agriculture, making it an effective tool to increase agricultural yield (Yan-e, 2011).

The acquisition of data and the extraction of agronomic knowledge on the spatial and temporal variability of crops can contribute significantly to the expansion of PA in sugarcane fields. Some studies have demonstrated the influence of soil attributes on sugarcane yield (spatially monitored). Using a yield monitor and decision trees algorithms, Souza et al. (2010) found that potassium and altitude were the most important attributes determining high yields. Cerri and Magalhães (2012) evaluated the correlations between sugarcane yield and some chemical and physical soil attributes. These correlations were found to be generally weak (< 0.5), and the authors concluded that a simple correlation is not sufficient to explain the spatial variability in yield, suggesting that characteristics other than soil attributes should be analysed. Working with sugarcane yield mapping, soil fertility attributes and attributes of sugarcane quality over 3 years, Johnson and Richard Jr. (2005) obtained non-significant, low and moderate correlations using linear Pearson correlations, suggesting that future studies should verify the influence of micronutrients on crop quality and yield. Rodrigues et al. (2013) also did not find patterns in the temporal stability of sugarcane quality parameters, suggesting that more crop cycles should be included in future assessments. Although few studies in the literature have reported on using of yield monitors in sugarcane fields to investigate the causes of spatial and temporal variability, some plot-scale studies have addressed the influence of soil attributes on yield (Bordonal et al., 2017; Rossi Neto et al., 2017; Dias et al., 1999).

As suggested by Cerri and Magalhães (2012), a simple Pearson's correlation between soil and plant parameters is not enough to explain yield. Advancements in data science and big data (Wolfert et al., 2017) may be able to address this bottleneck. Some studies reported in the literature have used data mining techniques, such as random forest (RF) algorithms (Breiman, 2001), to estimate sugarcane yields (Everingham et al., 2016; Bocca and Rodrigues, 2016; Bocca et al., 2015), showing the potential of these tools. RF methods have been widely adopted for certain agricultural problems, such as remote sensing analysis (Lebourgeois et al., 2017; Parente et al., 2017), leaf nitrogen levels (Abdel-Rahman and Ahmed, 2008) and classifying sugarcane varieties (Everingham et al., 2007). RF were used in many problems of yield estimation (Park et al., 2005; Tulbure et al., 2012; Fukuda et al., 2013; Newlands et al., 2014; Jeong et al., 2016), particularly in sugarcane fields (Everingham et al., 2009; Everingham et al., 2015a; Everingham et al., 2015b; Everingham et al., 2016). RF algorithms can handle large volumes of data, use categorical variables as predictors, measure the degree of importance of the predictive variables, and output the class probability and is robust against overfitting, even for slightly imbalanced datasets (Khoshgoftaar et al., 2007). Although previous studies have addressed sugarcane crops, none of them use yield monitor data spanning multiple years to assess the influence of soil attributes on sugarcane yield.

In this context, the objective of this paper was to investigate the relationship between physical and chemical soil attributes and sugarcane yield in order to identify the determinant parameters that define the spatial and temporal variability of yields. Thus, we used the computational environment created to support agricultural research, data acquisition, data formatting, data verification, data storage and analysis that was described in Driemeier et al. (2016) and developed with the objective of assisting PA studies. From large volumes of data obtained through soil and plant monitoring, it is possible to obtain new and relevant agronomic knowledge that can help producers increase yields and production profitability, thereby increasing the efficiency and sustainability of the sugarcane industry based on site-specific crop management.

2. Materials and methods

The data used in this paper are derived from two experimental sugarcane fields used for PA projects and are stored in the Agricultural Database (BDAgro) of the Brazilian Bioethanol Science and Technology Laboratory (CTBE/CNPEM). The first experimental area, with an area of 30 ha, is located at the Pedra Mill (PeM - São Paulo - Brazil - 21°16′36.94″S, 47°18′31.31″W - 583 m), and the second, with an area of 10 ha, is located at São João Mill (SJM - São Paulo - Brazil - 22°23′37.21″S, 47°18′31.31″W - 640 m). The slope of the areas is 10% and 2% for PeM and SJM fields, respectively. The sugarcane varieties, chosen according to the local climatic conditions and the soil type, were CTC09 and SP80-3280 for PeM and SJM, respectively. The full details and initial objectives of the PeM and SJM project experiments were reported by Magalhães et al. (2014) and Rodrigues et al. (2012), respectively. The main difference in the management of the two experimental fields is related to soil fertilization. At PeM, nitrogen (according to expected yield), phosphorus and potassium (according to the laboratory soil analyses) were applied at variable rates throughout the entire crop cycle (3 crop seasons), while at SJM, fertilizers were not applied during the experimental period (2 crop seasons). The areas were sampled in a regular grid of 50 × 50 m and 30 × 30 m for the PeM and SJM with a total of 107 and 117 soil sampling points, respectively (Fig. 1). Soil samples collected in the superficial layer (0.00 to 0.20 m) were submitted for wet-chemical laboratory analysis. The soil attributes assessed were soil organic matter (OM), pH, phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), hydrogen + aluminium (H + Al), the sum of bases (SB), cation exchange capacity (CEC) and base saturation (BS). PeM was evaluated during the crop seasons of 2012 to 2014, and SJM was evaluated in 2011 and 2012. The first year of sugarcane production (lasting from 12 to 18 months) is defined as the cane plant, and successive years (12 months) are defined as ratoon. In Brazil, the length of the average sugarcane cycle, from one planting to the next, is approximately five years. In the present experiment, the sugarcane crops corresponding to the evaluated years were the cane plant, 1st ratoon, and 2nd ratoon for PeM and the 2nd and 3rd ratoons for SJM. The experimental fields were harvested using a yield monitor coupled to the sugarcane harvest (SIMPROCANA®, ENALTA, São Carlos, Brazil).

2.1. Data analysis

The yield data were reduced to a soil sample grid by linear polynomial surface regression (*fittype function*) using Matlab software (MathWorks, Natick, Massachusetts) in the buffer zone (Fig. 1 – detail) according to the linearization method described by Driemeier et al. (2016). The soil chemical attributes were converted to logarithms of the concentrations. The logarithmic scale reduced the positive asymmetry of the distribution, which was both physically and chemically justifiable (Atkins and de Paula, 2010). The next step was to remove extreme values, which could cause detrimental biases for correlations, covariance, and subsequent analyses, from the datasets. Any input that