Computers and Electronics in Agriculture 130 (2016) 13–19

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

A computational environment to support research in sugarcane agriculture

Carlos Driemeier[a], Liu Yi Ling[a], Guilherme M. Sanches[a,b], Angélica O. Pontes[a], Paulo S. Graziano Magalhães[a,b,*], João E. Ferreira[c]

[a] Brazilian Bioethanol Science and Technology Laboratory (CTBE), National Center for Research in Energy and Materials (CNPEM), Caixa Postal 6192, CEP 13083-970 Campinas, São Paulo, Brazil
[b] School of Agriculture Engineering – FEAGRI, University of Campinas – UNICAMP, Campinas, SP, Brazil
[c] Institute of Mathematics and Statistics – IME, University of São Paulo – USP, São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form 23 August 2016
Accepted 4 October 2016

Keywords:
Precision agriculture
Sugarcane
Database
Workflow

ABSTRACT

Sugarcane is an important crop for tropical and sub-tropical countries. Like other crops, sugarcane agricultural research and practice is becoming increasingly data intensive, with several modeling frameworks developed to simulate biophysical processes in farming systems, all dependent on databases for accurate predictions of crop production. We developed a computational environment to support experiments in sugarcane agriculture and this article describes data acquisition, formatting, storage, and analysis. The potential to support creation of new agricultural knowledge is demonstrated through joint analysis of three experiments in sugarcane precision agriculture. Analysis of these case studies emphasizes spatial and temporal variations in soil attributes, sugarcane quality, and sugarcane yield. The developed computational framework will aid data-driven advances in sugarcane agricultural research.

1. Introduction

Sugarcane is an important crop mainly in tropical and sub-tropical countries. Brazil is the largest sugarcane producer, with 9 Mha cultivated to produce 659 million Mg of sugarcane in the 2015/2016 season, resulting in 34,600 Mg of sugar and 29 billion L of ethanol (CONAB, 2015). In addition to sugar and ethanol, Brazil is today the country with the largest installed capacity of biomass-based electricity generation (IRENA, 2015). In 2015, the supply of electricity from biomass had estimated growth of 7%, with a total generation over 22 TW h, where sugarcane accounts for 80%.

Several modeling frameworks such as AUSCANE, QCANE, APSIM, MOSICAS and CANEGRO (Marin et al., 2011) are increasingly being employed to simulate biophysical process in sugarcane farming systems. They are all dependent on databases, exemplifying the many ways in which agriculture is moving towards intensive data acquisition and processing. In addition, agriculture worldwide is witnessing a growing adoption of the so-called Precision Agriculture (PA), which comprises a set of tools to help farmers understand and manage soils and crops inherent spatial and temporal variability. PA relies on collection, analysis, processing, and synthesis of voluminous georeferenced data, which can be collected from a number of different technologies (Zamykal and Everingham, 2009). Research and technology on PA have advanced considerably in the past 20 years (Bramley, 2009). Due to its intense use of information, PA has grown and evolved to incorporate the best of multidisciplinary science and technology (Zamykal and Everingham, 2009), requiring farmers to look at their business from different perspectives (Srinivasan, 2006).

Sugarcane production system, however, differs substantially from major staple crops, affecting development and adoption of agricultural technologies. Comparison between a major cereal (*e.g.*, wheat) and sugarcane highlights some key differences. Wheat area worldwide is 215 Mha, primarily in temperate zones, compared to 26 Mha of sugarcane, primarily in tropical developing countries, especially in Brazil. Furthermore, the harvested part of cereal crops is the grain, with mean yields of 3.2 Mg ha^{-1} for wheat, compared to 71 Mg ha^{-1} for harvesting the stalks of sugarcane in 2013 (FAO, 2015). Differences in area and location make sugarcane a small fraction of the global market for agricultural technology. In addition, the high tonnage of sugarcane requires dedicated

* Corresponding author at: Brazilian Bioethanol Science and Technology Laboratory (CTBE), National Center for Research in Energy and Materials (CNPEM), Caixa Postal 6192, CEP 13083-970 Campinas, São Paulo, Brazil.
E-mail address: paulo.graziano@bioetanol.org.br (Paulo S. Graziano Magalhães).

http://dx.doi.org/10.1016/j.compag.2016.10.002
0168-1699/

technologies, such as tailored yield monitors (Magalhães and Cerri, 2007). Due to specificities of the sugarcane system, and despite rapid adoption of auto-steer in tractors and harvesters (Bramley and Trengove, 2013; Silva et al., 2011), PA is not yet adopted by the sugarcane-based sugar-ethanol industry as it is for other agricultural systems (Gebbers and Adamchuk, 2010). According to surveys conducted in Brazil (Anselmi et al., 2014; Avanzi et al., 2014; Silva et al., 2011) and Australia (Bramley and Trengove, 2013), low PA adoption can be explained by four factors: relative advantage (usefulness), compatibility, trialability and observability. For sugarcane production, perceived usefulness is correlated with increased crop yield, reduced costs, and improved management. On the other hand, high costs of equipment, lack of qualified staff and lack of information on PA technologies were pointed by sugarcane farmers as the main barriers (Silva et al., 2011).

In this context, efforts have been primarily dedicated to experiments aiming at establishing the scientific grounds and demonstrating the advantages of PA techniques applied to sugarcane (Portz et al., 2011; Rodrigues et al., 2012). Because of these goals, characterization of soil and plant attributes in experiments is much more comprehensive than the expected for large-scale PA practice. Furthermore, testing data acquisition technologies and contextualizing their outputs are important goals of the experimentation stage. Considering the above, treating the diversity of measurable attributes is a critical point in experimentation for sugarcane PA.

The data-driven character of PA has attracted the attention of the research community from many different areas. For instance, there are studies on clustering algorithm to delineate management zones (Tagarakis et al., 2012), data acquisition techniques with remote sensing (Mulla, 2013; Song et al., 2009), and software architecture for data analysis and integration of sensor based PA monitoring (Chen et al., 2015).

In this work, we present a computational environment created to support sugarcane agricultural research, including but not limited to research in PA. Data acquisition, formatting, verification, storage, and analysis are discussed. To demonstrate the applicability of the computational environment, data of soil chemistry, sugarcane quality, and sugarcane yield from three experiments are jointly analyzed and discussed.

2. Computational environment

2.1. Handling of raw data

Sugarcane agricultural experiments may include several sources of raw data, including data acquired by different analytical laboratories and by various types of sensors. The current version of the computational environment is able to process data in matrix formats. Processing of images (*e.g.*, from unmanned aerial vehicles and satellite) is foreseen as a future upgrade for the system.

The database has an expandable set of allowed matrix formats – essentially one matrix format for each type of measurement. To assure that matrixes are properly recorded in the database, we routinely handle raw data following the tasks presented in Fig. 1.

Using spreadsheets, raw data from sensors and laboratory files are converted into data matrixes consistent with the predefined database formats. Such data matrixes are verified and then inserted into the database. Importantly, data acquisition and formatting are performed by agricultural field scientist, while verification and insertion are performed by computational workers. Among other advantages, this division of tasks assures an independent verification of data veracity. Verifications include matrix formats, measurement units, and typical range of values acceptable for a certain measured attribute. Once verified, data matrixes are inserted into the database using python-generated SQL scripts.

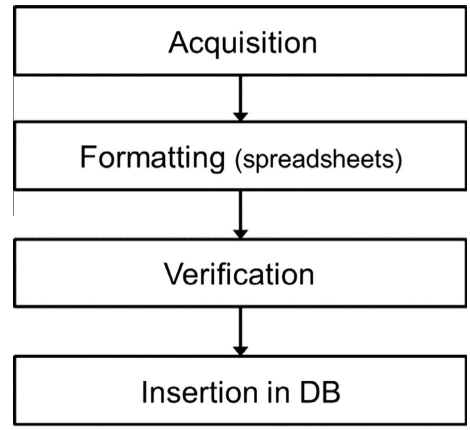

Fig. 1. Tasks for handling of raw data prior to insertion into the database.

2.2. Database

A relational database for sugarcane agricultural experiments was created and named *BDAgro – CTBE Database of Agricultural Experiments*, as detailed in a Technical Memorandum (Pontes et al., 2014). BDAgro was constructed having *PostgreSQL* as relational database management system and *pgAdmin* as database administration and development platform.

BDAgro conceptual model, *i.e.* entity-relationship model (Elmasri and Navathe, 2010), includes entities associated to management and responsibilities (*e.g.*, records of projects and responsible persons). Nevertheless, more relevant for the analytical purposes of the computational environment, BDAgro represents agricultural experiments through the following entities:

- Experiment is defined by a certain land area during a certain period of time. The land area is most often an open agricultural field, but may also be inside close environments such as greenhouses.
- Event is one important fact within one experiment. Events may be of three types: (i) *intervention*, associated with change in experimental land area (*e.g.*, harvest, soil fertilization); (ii) *characterization*, associated with data acquisition without change in land area (*e.g.*, characterization of soil attributes); and (iii) *planning*, representing a record associated with neither physical change in land area nor new data acquisition (*e.g.*, nutrient prescription).
- Static data is data generated by events. It is termed *static* because each event is defined at a specific moment within one experiment. Static data has x and y spatial coordinates as attributes. Additional attributes depend on type of static data (*i.e.*, on type of measurements). Soil attributes, sugarcane quality, and sugarcane yield are examples of types of static data.
- Dynamic data is data acquired continuously during the course of one experiment. Meteorological information is one example of dynamic data.

We will refer to these entities as the article follows. However, the analysis of the case studies does not yet include any *dynamic data* because of the complexity of agricultural analysis using fine temporal granularity.

2.3. Data analysis

We adopted the Work-Event-Data-flow (WED-flow) approach (Ferreira et al., 2010) as the methodology for modeling analysis