Review Article

Assessment of data fusion oriented on data mining approaches to enhance precision agriculture practices aimed at increase of Durum Wheat (*Triticum turgidum* L. var. *durum*) yield

A. D’Accolti, S. Maggio, A. Massaro*, A. M. Galiano, V. Birardi, L. Pellicani
Dyrecta Lab srl, Research Institute, via Vescovo Simplicio 45, 70014 Conversano (BA), Italy

(Received: 11-09-2018; Revised 08-11-2018; Accepted 18-11-2018; Published Online 17-12-2018)
*Corresponding author (E-mail: alessandro.massaro@dyrecta.com)

Abstract

In 2050, world population will reach a total of 9 billion inhabitants and their food demand have to be satisfied. Durum wheat (*Triticum turgidum* L. var. *durum*) is one of the most important food crop and its consumption is increasing worldwide. Productivity growth in agriculture and profitable returns are strongly influenced by investment in research and development, where Precision Agriculture (PA) represents an innovative way to manage farms by introducing the Information and Communication Technology (ICT) into the production process. It is known that farms activities produce large amounts of data. Today ICT allows, with electronic and software systems, to collect and transfer automatically these data, thus increasing yields and profits. In this direction significant data are processed from agricultural production, and retrieved to extract useful information, important to increase the knowledge base. Data from multiple data sources can be processed by a Data Fusion (DF) approach, able to combine multiple data sources into a unique database system. Raw data are transformed into useful information, thus DF improves pattern recognition, analysis of growth factors, and relationship between crops and environments. Data Fusion is synonym of Data Integration, Sensor Fusion, and Image Fusion. By means of Data Mining (DM) it is possible to extract useful information from data of the production processes thus providing new outputs concerning product quality and “health status”. The following literature take into account the DF and DM techniques applied to Precision Agriculture (PA) and to cultivation inputs (water, nitrogen, etc.) management. We report also last advances of DF and DM in modern agriculture and in precision durum wheat production.

Keywords: Data science; Crop Cultivation; Predictive Models

Introduction

FAO (FAO, 2009) declared the world population in 2050 will reach a total of 9 billion inhabitants and food demand will increase. In developed countries, 80% of the increase in agricultural production will have to come from higher yields and intensification of cultivation, and only 20% from the expansion of arable land (Tilman et al., 2011; Grafton et al., 2015; Hunter et al., 2017; Conijn et al., 2018). The major agriculture role is to provide the food without harming environment and by avoiding inputs dispersions. Nowadays a farmer, together with crops, collects an increasing amount of data produced by satellite navigation systems, sensing, and monitoring technology. An Information and Communication Technology (ICT) enabled-service can provide data storage and processing to improve the production processes in the agricultural sector (George et al., 2011; Salampasis and Theodoridis, 2013). This combination of technology, agriculture, and data acquisition, has been defined as Precision Agriculture (PA). The objectives of PA are the advances in genetics, agricultural practices, weather forecasting, farm management aimed at yield optimization, detection of the field variability, and decision-making processors ‘Decision Support Systems’ (DSS). Agricultural practices implemented by means of PA, include site-specific prediction and distribution, science mapping approach, disease identification, crop nitrogen and water use efficiency (Whelan and Taylor, 2013; Pedersen and Lind, 2017; Zhang, 2017). Measurement and the knowledge of the field variations are the first steps in the PA approach, assessing the spatial and temporal variability of plant growth rate and crop yield (Bobryk et al., 2018).

The developments in remote and proximal sensors (Lamb and Brown, 2001; Adamchuk et al., 2004; Oliver et al., 2013; Mulla, 2013; Mohd Kassim et al., 2014) provided new data sorts and sets. The employed remote sensors are related to geo location by global positioning system (GPS), and by global navigation satellite system (GNSS) technologies (Guo et al., 2018) e.g. to individuate yield distribution (Stafford, 2000; Shannon et al., 2018). The raw data processing, collection and organization follow scientific criteria to the creation of information systems. These latter are modeled after agricultural sector information needs which the most important is the control about decision-making processes (Milovic and Radojevic, 2015). Farmers, thanks to information systems can empower the control of their resources. Furthermore, agricultural production, processing, trading, and marketing improve with the technological progress as ICT (Vidanapathirana, 2012). The Data Fusion (DF) it has been applied by researchers in the field of satellite image processing for agricultural purposes (Rodrigues et al., 2009). Kussul et al. (2015) and Jia et al. (2016) proving the possibility of subjecting images to DF in a process called Image Fusion (IF). Data Mining (DM) algorithms are applied...
in agriculture to identify hidden associations and patterns (agricultural patterns), make predictions or take decisions in multiple area (Kaur and Singh, 2014). DM includes disciplines such as statistics, computer science and artificial intelligence (Alpaydin, 2014).

Durum wheat (Triticum turgidum L. var. durum) is one of the most cultivated wheat in the world. It’s cultivated in semiarid regions such as the Middle East, the North American Great Plains, Mediterranean Europe, and North Africa. Durum wheat represents 8 to 10% of all the wheat cultivated area and therefore, is considered as a minor crop compared to bread wheat, but its consumption trend is worldwide increasing. Durum wheat end products such as pasta, couscous, bulgur, frekeh, flat or leavened typical breads are staple food especially for people from Northern Africa, the Middle East and Mediterranean Europe, but they’re also appreciated from all over the world (Elias, 1995). Genotypes, environments (weather and nutrition) and crop management affect durum wheat end products quality. The main qualitative features of durum wheat would be related to different supply chains, for example protein and gluten content are most important for ‘Pasta Industry’, while grain moisture and impurity are basal for Milling Industry. Grain yield is the main goal for the farmer, so breeding programs are all projected towards obtaining varieties with yield stability, resistance to biotic and abiotic stresses and adaptation to various environments. Majors factors causing decrease in yield and grain quality are drought, high temperature, and biological stress (terminal stress) in the ripening phase (Troccoli et al., 2000).

**Data fusion**

The DF consists of data derived from multiple resource, associated into systems and models. The agricultural enterprises employ useful information obtained with DF e.g. to help farmers in taking decision, such as choosing the right crop for the right soil and the right environment. DF is related to similar concepts as Data Integration (DI), Sensor Fusion (SF), and Image Fusion (IF). According to Malviya et al. (2015) the crops productivity increased by means of data processing and predictions carried out by DF systems.

**Data integration**

The DI involves the combination of data coming from different data sources such as soil databases, long-term data on carbon balance across different climate zones and vegetative land covers, digital elevation models, regional and national inventories, remote sensing data, geophysical data, socio-economic, and many other data sets. The latest agricultural application of DI has been described by Nabrzyski et al., (2014) and Bruce and Reynolds (2016). The soils characterization is an example of DI application (Mouazen et al., 2016) by using VIS-NIR-SWIR spectroscopic configurations (Rosero-Vlasova et al., 2016), and proximal sensors (Cho et al., 2016; Veum et al., 2017) to describe more than one soil of interest (Mahmood et al., 2012).

**Sensor fusion**

Data sources are typically data sensors. Sensors detect, measure, store, process and communicate the state of the ambient inside and/or outside of a living body and its variations in time and space dimension (D’Amico et al., 2015). The diagnostic-investigative science involving sensors to monitor qualitatively and quantitatively environments and objects is called sensor science. ‘Remote sensing’ involves observation and measurement of specific object features or target from long distance. Besides ‘Proximal sensing’ consists of the monitoring activity working in a close range or in contact with the area of measure (Proffitt et al., 2006; Kumar and Ilango, 2018). Sensor Fusion (SF) is a specialized technique of DF in agreement with information merging from two or more sensors (Klein, 2004; Gustafsson, 2010; Ji et al., 2017). Finally, other researchers (Riezzo, 2013; Todorovic, 2013) studied the integration of different sensors able to control the irrigation by means of a Decision Support System - elaborating all sensor data.

**Image fusion**

The IF process (Pohl and van Genderen, 2014, 2015)combines two or more registered images of an identical scene into a more interpretable single one. In a review of ‘Remote Sensing’ IF methods, Ghassemian (2016) combined all the methods to obtain an image, having the best characteristics of both spatial and spectral resolution. So, for the full exploitation of multisource data, advanced analytical or numerical image fusion techniques have been developed. The goals of the IF are the improvement of spatial resolution and classification accuracy, the enhancement of display features capabilities, the geometric precision, and the replacing or the repairing defects of image data. This method is particularly relevant in heterogeneous environments as cloud prone-landscapes (Knauer et al., 2016).

**Data mining**

Data Mining (DM) is defined by Ramesh and Vardhan (2013) as the process of extracting useful information from large datasets. Modern DM techniques establish relationships and associations rules of different observations sources. According to many authors (Patel and Patel, 2014; Mistry et al., 2016; Kodeeshwari and Ilakkiya, 2017) DM techniques were used mainly for classification and clustering, but can perform association, regression, and predictive analysis. Outputs of crop yield (Manjula and Narsimha, 2016) performed by DM engines have been used by farmers to improve crop performance (Chouhan, 2016; Jiménez et al., 2016). The book written by Mucherino et al. (2009) is a broad overview of recent DM techniques and applications in agriculture and is the first completely dedicated to emerging research fields.

**Precision Agriculture**

Appliance of data fusion and data mining to delineate management zones

The estimation of soil and crop variability (Cavallo et al., 2016) are main topics of PA process requiring a deep knowledge (De Benedetto et al., 2013). The ‘Management Zones’ (MZs) are intra-field homogeneous areas with similar
influences on yield characteristics (Sissons et al., 2016). The data sources processed to obtain MZs are multi-year yield analysis, soil survey public, remote sensing, topography/landscape, soil properties, and grower knowledge-based. The delineation of MZs has been adopted in site-specific management (Navar et al., 2017). The DF techniques used to elaborate MZs are based on proximal (Mahmood, 2013; Castrignanò et al., 2015; Rodrigues et al., 2015) and remote sensing (De Benedetto et al., 2013; Xie et al., 2013; Gevaert et al., 2015). Authors (Fraisse et al., 2001; Mzuka et al., 2005; Li et al., 2007; Pedroso et al., 2010; Moral et al., 2010; Guasaferrro et al., 2010; Davatgar et al., 2012; Tagarakis et al., 2013; Pantazi et al., 2015; Shaddad et al., 2016; Buttafuoco et al., 2017; Castrignanò et al., 2017; Schenatto et al., 2017; Servadio et al., 2017; Castrignanò et al., 2018; Georgi et al., 2017) described the most widely processes used to obtain MZs. Another research (Schemberger et al., 2017) described the DM algorithms suitable to delineate the MZs. The MZs benefit farmer after yield increasing and contribute to reduce nitrogen losses compare to conventional nitrogen management (Khosla et al., 2002; Koch et al., 2004)

Yield prediction

Yields estimation and prediction are pivotal for food safety and company decision-making processes (Chaudhari et al., 2010). The forecasting productivity models allow to obtain updated information regarding the internal product demand, the availability for export, the market data. In the grain production systems, yield information is used to optimize the cultivated areas, improve cultivation techniques and boost yields. Historically, production inputs have been managed according to “recommendation domains” as reported by Jiménez et al. (2016). Nowadays it is possible to optimize resources by using agro ecosystem mathematical quantifications (Luschei, 2001; Wagner, 2004). Agronomic indicators such as phenological and vegetative indices typically characterize crop models and estimate yields. The morpho-physiological traits of the crop, the weather variables, the knowledge of the farmer, the cultivation conditions, the use of sensors and spectroradiometric features of the culture, are inputs for predictive models (Lamba and Dhaka, 2014). Hyperspectral data from spectroradiometers have been applied to estimate wheat features such as nitrogen content, water content, and crop yield (Thorpe et al., 2017) or yields maps from satellite data (Zheng et al., 2016) subjected to IF. Data extracted from a growth model and a radiative model have been coupled and fused by Zhang et al. (2016), on the basis of vegetative indices and culture management parameters such as sowing date, sowing rate, and nitrogen rate. This DF model shown accuracy and efficiency to predict crop yield on a regional scale.

Jambekar and Saquib (2018) hypothesized grain yield prediction by using DM techniques. They processed weather data from 1950 to 2013, total area dimension, average temperatures, amount of precipitation, area under irrigation, and annual yield. As assessed by authors, Multiple Linear Regression, Random Forest Regression, and Support Vector Regression are most used DM algorithms to predict grain yield. In season wheat yield prediction has been performed using crop simulation model constituted by Geographic Information System (GIS), remote sensing and ground observed data (Chaudhari et al., 2010). Verma et al. (2018) proposed a model to help farmers forecasting yields. They successfully integrated and fused data by clustering with Fuzzy C Means (is a method of clustering which allows one piece of data to belong to two or more clusters) and by classifying by neural networks considering variable as biomass, temperature, rainfall, and solar radiation data. Pantazi et al. (2016) tested ‘Artificial neural networks’ (ANN) such as Supervised Kohonen Networks (SKN), counter-propagation artificial neural networks (CP-ANN) and XY-fusion (XY-F) oriented on yield prediction for a single cropping season to understand yield limiting factors. The physical soil parameters were obtained, with a visible and near-infrared spectroscopic sensor (VIS-NIR) fused with crop growth indices derived from a satellite. The authors associated high definition soil and crop data with classes of is frequency referring to yield. One of the tasks entrusted to DM techniques is the yield prediction based on available data. SKN algorithm shown the best validity and precision to predict wheat yield.

Precision Durum Wheat Production

We reported the latest literature within the precision durum wheat production.

Nitrogen management

The wheat productivity and quality are mostly influenced by climate variables, nitrogen supply (Morari et al., 2013; Tedone et al., 2018) and land characteristics. The authors (Cossani and Sadras, 2018) consider the concept of co-limitation in water and Nitrogen availability to explain cereal yield gaps. Buttafuoco et al. (2017) described a geo statistical approach to delineate the MZs and set up a site-specific management in a durum wheat field in Southern Italy. Servadio et al. (2017) delineated MZs to exert VRT in a durum wheat field in Central Italy. The authors used a combine harvester equipped with grain mass flow sensor, GPS, and Precision Land Management Software to collect data and investigated the soil geo referenced physical-chemical properties, such as structural stability, water content, shear strength, and total nitrogen; data were processed with two cluster analyses applying a fuzzy algorithm. Morari et al. (2018) proposed Variable Rate Fertilization (VRF) and precision harvesting to optimize durum wheat cultivation in Northern Italy. The VRF mitigated the weather impact that can afflict negatively Nitrogen Use Efficiency (NUE). Main effects of these PA practices are the reduction in environmental impact and increase in grain content of gluten protein, thus improving high quality. Basso et al. (2016) assessed durum wheat to VRF response using remote sensing.

Yield estimation

Some researchers (Stewart et al., 2002; Cavallo et al., 2016) used geographical data for assessing soil variation in a durum wheat field association. They demonstrated the relationship between structural properties of the soil and durum wheat yield (Cavallo et al., 2016), and others
(Aparicio et al., 2000; Inurreta-Aguirre et al., 2018) considered phenological classifications and vegetative indices such as LAI and NDVI to estimate durum wheat yield in a humid Mediterranean climate.

**Data science applied to durum wheat production**

The last-two-years literature research produced the following two main results as regards DF and DM applied to durum wheat farming:

- The DF algorithm based on the Kalman filter to estimate leaf area index evolution, in durum wheat by using field measurements and Moderate Resolution Imaging Spectroradiometer surface reflectance data (Novelli et al., 2016).

- The estimation of durum wheat growth, nitrogen status, and grain yield by hyperspectral data mining. The authors compared the spectral components to estimate the durum wheat traits, and developed a genetic algorithm to identify the most relevant spectral features; grain yield has been optimally estimated from canopy spectral measure mensusing the genetic algorithm approach (Thorpe et al., 2017).

**Conclusion**

Data science changes farming approaches from empirical to measurable. In agriculture the ICT introduction opens to new challenges as well as the development of the specific-site systems and the yield estimation. The prescription and fertilization maps improve the reliability of yields prediction and historicizing, the soils feature data, the growth and development models, and the seasonal forecasts. The data collection and data processing are performed by field sensors and complex algorithms, making available low-cost equipment of PA for farmers. This new working way in agriculture is the key to increase and improve production, with a view to sustainability, traceability and adaptability to climate change. The DF approach combines multiple data sources to obtain best outputs. The application of DF techniques has a large extent and needs more research work. The literature analyses highlight the scarcity of DF and DM practices applied to durum wheat cultivation. For this reason, in this short review, we referred to other crops and especially to winter wheat when not explicitly stated in the text. It is necessary in the future research to carry out more studies related to durum wheat because of the different cultivation area, the various climatic conditions, the different biological and genetic traits, dissimilar grain and end products, and the multiple qualitative features that characterize it when compared to winter wheat. The PA practices, for durum wheat cultivation, change according to the size and location of the farm. By taking into account a national case study, durum wheat cultivation areas, specially in Southern Italy, are located in environments characterized by pedological, orographic, and climatic high in homogeneity. Physical environment variability is present during all durum wheat cycle showing high differentiation in both cultivation areas, with limited extensions as well as fields large few hectares. In this context site-specific, according to the needs distribution of cultivation inputs influences positively economic and environmental sustainability. This new agriculture is the key to improve yield, traceability, and adaptability to climate change.

**Acknowledgments**

This research was supported by Santacroce Giovanni S.P.A. in the framework of the project: Data fusion platform oriented to data mining processing of durum wheat production chain data: “Wheat Data Fusion” [Piattaforma di data fusion orientata al data mining processing dei dati della filiera del grano duro: “Wheat Data Fusion”]. Authors thanks for their contributions the following researchers of Dyrecta Lab: G. Birardi, V. Calati, V. Custodero, F. De Carlo, G. Fanelli, A. Leogrande, G. Lonigro, N. Malfettone, L. Pellicani, O. Rizzo, N. Savino, M.M. Sorbo, D. Suma, and F. Tarulli. Authors also gratefully thanks for their support A. Bruno, C. Di Iorio, A. Giansante, R. Marinaccio, S. Medei, N.P. Santacroce, M. Prencipe, A. Saldarella, S. Saldarella, A. Tomaiuoli in the realization of this work.

**References**


Jimenez, D., Dorado, H., Cock, J., Prager, S.D., Delerce, S., Grillon, A., Bejarano, M.A., Benavides, H., and Jarvis,


https://www.phoenixpub.org/journals/index.php/jfna


