Proceedings of the First International Conference on Agro Big Data and Decision Support Systems in Agriculture

Universitat de Lleida

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

September 27-29 2017, Montevideo, Uruguay

With the support of

Welcome

Dear Colleagues and Friends,

Welcome to Montevideo to the 1st International Conference on Agro Big Data and Decision Support Systems in Agriculture. It is our great pleasure to have you here and would like to wish you a pleasant stay in this beautiful city and a fruitful conference.

The conference has been organized jointly by the Engineering School of the Universidad de la República, Uruguay, hosting the event, and by the CYTED-funded BigDDSAgro network (Red Iberoamericana de Agro-Bigdata y Decision Support Systems para un sector agropecuario sostenible). The conference has also received the support of the EURO Working group of Operational Research in Agriculture and Forest Management.

The BigDSSAgro network started its activities in 2016, and it comprises research groups from 39 institutions (universities, public research centers, and private companies), coming from 15 different countries in Latin America, Europe and Oceania. The main objectives of the network are promoting the interaction, the cooperation and the transfer of knowledge and technologies related to heterogeneous information systems supporting decision making in agriculture. Among other activities, the BigDSSAgro network organizes regular meetings among its participants, and develops student and researcher exchanges, as well as supporting the proposal of joint research and cooperation activities. In 2016, during the first general meeting held at Santiago de Chile, the participants of the network decided to support the organization of an international conference, open to all researchers in the area, devoted to big data and decision support systems in agriculture, which combine models, databases and algorithms to support many real-world decision-making problems. It was decided that the first edition of this event would be held in Montevideo, Uruguay.

To develop this initiative, a program committee was integrated with researchers from all the world, widening the appeal of the event and receiving a support well exceeding the groups participating in the network. A CFP was prepared and circulated in different communities, newsletters and discussion groups, and the papers received in answer to the call were subject to a peer-review process, to guarantee the quality of the results. Also, two special issues were agreed with prestigious journals in the area, ”Computers and Electronics in Agriculture” and ”Annals of Operations Research”, opening an outlet for publication of extended versions of papers presented at the conference.

We are proud to report that the effort was very successful, leading to a conference program comprising more than 50 oral presentations and 5 poster presentations. The conference also includes two keynote presentations, by Dr. Pascale Zaraté from Université de Toulouse-Institut de Recherche en Informatique de Toulouse, France, and by Dr. Walter Rossing from Wageningen University, The Netherlands, plus a seminar given by Dr. Emilio Carrizosa from University of Sevilla, Spain. In parallel to the academic tracks, the conference also includes an event for industry and government presentations on ICT in Agriculture, including 12 contributed presentations, 3 invited presentations, and two keynotes by Dr. Walter Baethgen from Columbia University, USA, and by Dr. Carlos Meira, from EMBRAPA, Brazil.

We hope this first edition marks the start of a very successful series of events, leading to improved cooperation in the community, which is very diverse as it includes researchers from agriculture, computer science, operations research, statistics, and other backgrounds, and which is facing an increasing demand for both conceptual proposals and for practical applications leading to improved efficacy and efficiency and to new productive and societal opportunities.

Héctor Cancela, Antonio Mauttone, Adela Pagès and Lluis Miquel Plà
Chairs BigDSSAgro 2017

Big DSS Agro Committee

General Chairs

Héctor Cancela, Universidad de la República, Uruguay
Lluis Miquel Plà-Aragonés, Universitat de Lleida, Spain

Program Chairs

Antonio Mauttone, Universidad de la República, Uruguay
Adela Pagès-Bernaus, Universitat de Lleida, Spain

Technical Program Committee

Renzo Akkerman	Wageningen University	The Netherlands
Víctor Albornoz	Universidad Técnica Federico Santa María	Chile
Luiz H. Antunes Rodrigues	Unicamp	Brazil
Walter E. Baethgen	Columbia University	USA
Carlos Bouza	Universidad de La Habana	Cuba
Pablo Chilibroste	Universidad de la República	Uruguay
Ángel Cobo	Universidad de Cantabria	Spain
Santiago Dogliotti	Universidad de la República	Uruguay
Fernando Garagorry	Empresa Brasileira de Pesquisa Agropecuária	Brazil
Frédérick Garcia	DigitAg	France
Marcela González Araya	Universidad de Talca	Chile
Andrew Higgins	CSIRO	Australia
Anders R. Kristensen	University of Copenhagen	Denmark
Concepción Maroto	Universitat Politecnica de Valencia	Spain
Lorena Prádenas	Universidad de Concepción	Chile
Daniel Rodríguez-Bocca	University of Queensland	Australia
Sara Rodríguez	Universidad Autónoma de Nuevo León	Mexico
Pablo Rodriguez-Bocca	Universidad de la República	Uruguay
Lorena Rodríguez-Gallego	Universidad de la República	Uruguay
Walter Rossing	Wageningen University	The Netherlands
Sérgio Serra da Cruz	Universidade Federal Rural do Rio de Janeiro	Brazil
Christian von Lücken	Universidad Nacional de Asunción	Paraguay
Andrés Weintraub	Universidad de Chile	Chile
Pascale Zaraté	Institut de Recherche en Informatique de Toulouse	France

Local Organizing Committee

Jorge Corral	Universidad de la República	Uruguay
Gastón Notte	Universidad de la República	Uruguay
Martín Pedemonte	Universidad de la República	Uruguay
Pablo Rodríguez-Bocca	Universidad de la República	Uruguay
Omar Viera	Universidad de la República	Uruguay

BigDSSAgro 2017

I International Conference on Agro BigData and Decision Support Systems in Agriculture
September 27 - 29, 2017, Montevideo, Uruguay

Program

Wednesday 27	Track A held at Room B21	Track B held at Room B22
9:00 – 9:30	*Registration*	
09:30 – 10:30	Wed.A1 : Remote sensing I	Wed.B1 : Artificial intelligence I
10:30 – 11:00	*Coffee Break*	
11:00 – 11:30	Opening (Amphitheater)	
11:30 – 12:30	Plenary Talk 1 : How to support Cooperative Decision Making? - Pascale Zaraté (Amphitheater)	
12:30 – 14:30	*Lunch Time*	
14:30 – 15:30	Wed.A2 : Optimization and simulation I	Wed.B2 : Atificial intelligence II
15:30 – 16:00	*Coffee Break*	
16:00 – 17:00	Wed.A3 : Optimization and simulation II	Wed.B3 : Data mining and bussiness intelligence I
17:00 – 18:30	Seminar of Data Science (Session 1) - Emilio Carrizosa (Room B21)	

Thursday 28	Track A held at Room B21	Track B held at Room B22
09:10 – 10:30	Thu.A1 : Remote sensing II	Thu.B1 : Decision support systems I
10:30 – 11:00	*Coffee Break*	
11:00 – 11:30	Poster Session	
11:30 – 12:30	Plenary Talk 2 : Model-aided learning for ecological intensification in agriculture - Walter Rossing (Amphitheater)	
12:30 - 14:30	*Lunch Time*	
14:30 - 15:30	Thu.A2 : Optimization and simulation III	Thu.B2 : Atificial intelligence III
15:30 - 16:00	*Coffee Break*	
16:00 - 17:00	Thu.A3 : Data envelopment and multicriteria analysis	Thu.B3 : Data mining and bussiness intelligence II
17:00 – 18:30	Seminar of Data Science (Session 2) - Emilio Carrizosa (Room B21)	

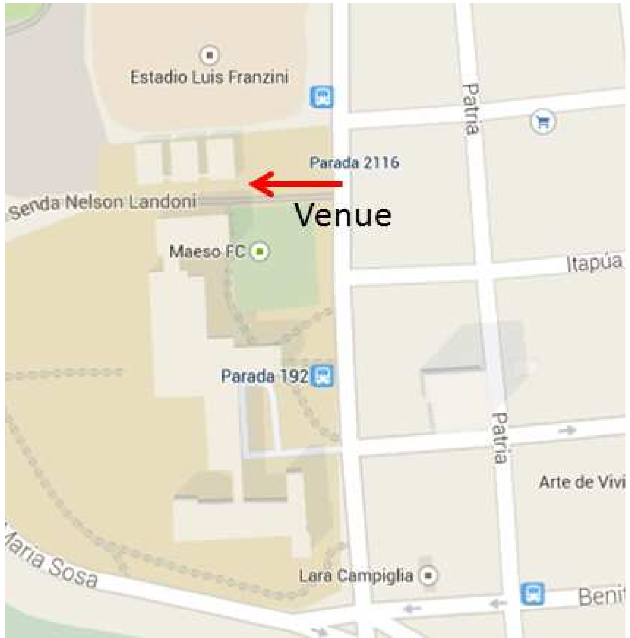
Friday 29	Track A held at Room B21	Track B held at Room B22
09:00 – 10:40	Fri.A1 : Descriptive mathematical models	Fri.B1 : Tecnhological developments
10:40 – 11:10	*Coffee Break*	
11:10 – 12:10	Fri.A2 : Optimization and simulation IV	Fri.B2 : Decision support systems II
12:10	Closure: Round table on future activities and cooperation opportunities	
	Free	
17:00 – 18:30	Seminar of Data Science (Session 3) - Emilio Carrizosa (Room B21)	

The First International Conference on Agro BigData and Decision Support Systems in Agriculture will be held at the Edificio Polifuncional "José Luis Massera"

The Registration and Information Desk is located in the module C of the building.

Address:
Senda Nelson Landoni c/ Julio Herrera y Reissig.
Montevideo, Uruguay.

Website:
http://www.bigdssagro.udl.cat/?q=node/75&language=en



Wednesday 27

Wed.A1 : Remote sensing I	Wed.B1 : Artificial intelligence I
Wed.A1 : 09:30 – 09:50 Remote monitoring of physico-chemical water parameters in rural water sources *Jean Betilio Mirón Acevedo, Mauricio José Grande Cóbar, Francisco Eduardo Huguet Mendez and Mauricio Pohl*	Wed.B1 : 09:30 – 09:50 Large data volume parallel clustering for computational cost reduction in Self-Organized Networks training *Mariela Azul Gonzalez, Pablo Montini, Jorge Martínez Arca and Lucía Isabel Passoni*
Wed.A1 : 09:50 – 10:10 Remote Sensing of Algal Blooms in the Uruguay River Based on Multispectral Satellite Imaging and Field Data *José Lezama, Fernanda Maciel, Francisco Pedocchi and Pablo Musé*	Wed.B1 : 09:50 – 10:10 Detection of Faults in WSNs based on NMF *Jimmy Ludeña-Choez, Juan José Choquehuanca-Zevallos and Efrain Mayhua-López*
Wed.A1 : 10:10 – 10:30 Sensor Data Analysis and Sensor Management for Crop Monitoring *Raquel Sosa, Leonardo Steinfeld, Andres Vera, Maite Ibarburu, Javier Schandy and Fernando Silveira*	Wed.B1 : 10:10 – 10:30 Analysis and selection of areas through clustering techniques for the Agropolis formation in Santander Colombia *Leonardo Talero-Sarmiento, Edwin Garavito-Hernandez, Henry Lamos-Diaz and Daniel Martinez-Quesada*

Wed.A2 : Optimization and simulation I	Wed.B2 : Artificial intelligence II
Wed.A2 : 14:30 – 14:50 Metaheuristic algorithms for multi-objective optimization in dairy systems *Gastón Notte, Héctor Cancela, Pablo Chilibroste and Martín Pedemonte*	Wed.B2 : 14:30 – 14:50 Body condition estimation on cows from 3D images using Convolutional Neural Networks *Juan Rodríguez Alvarez, Mauricio Arroqui, Pablo Mangudo, Juan Toloza, Daniel Jatip, Juan Rodriguez, Alejandro Zunino, Claudio Machado and Cristian Mateos*
Wed.A2 : 14:50 – 15:10 Simulated Annealing in the Operational Forest Planning *Paulo Amaro V. H. Dos Santos, Arinei Carlos L. Da Silva and Julio Eduardo Arce*	Wed.B2 : 14:50 – 15:10 Computer vision based system for apple detection in crops *Mercedes Marzoa, Gonzalo Tejera and Matias Di Martino*
Wed.A2 : 15:10 – 15:30 Identifying trade-offs between sustainability dimensions in the supply chain of biodiesel in Colombia *Orjuela-Castro Javier Arturo, Aranda-Pinilla Johan A. and Moreno-Mantilla Carlos Eduardo*	Wed.B2 : 15:10 – 15:30 Barley yield prediction under different fertilization treatments using machine learning and UAV imager data *Sara Rodriguez, Manuel Jimenez, Rigoberto Vazquez, Hugo Escalante*
Wed.A3 : Optimization and simulation II	**Wed.B3 : Data mining and bussiness intelligence I**
Wed.A3 : 16:00 – 16:20 Dynamic diet formulation responsive to price changes: a feed mill perspective *Adela Pagès Bernaus, Lluís Miquel Plà Aragonés, Jordi Mateo Fornes and Francesc Solsona*	Wed.B3 : 16:00 – 16:20 Analysis of decomposition parameters of green manure in the Brazilian Northeast with Association Rules Networks *Dario Calçada, Solange Rezende and Mauro Teodoro*
Wed.A3 : 16:20 – 16:40 Mathematical modeling under uncertainty for supply chain of sugar cane in Cuba *Esteban López Milán and Lluis Miquel Plà Aragonés*	Wed.B3 : 16:20 – 16:40 Dealing with derivatives for water quality management *Sira M. Allende Alonso, Carlos N. Bouza-Herrera and Rajesh Singh*
Wed.A3 : 16:40 – 17:00 Planning tool for the multisite pig production system based on stochastic optimization *Esteve Nadal and Lluis M. Pla*	Wed.B3 : 16:40 – 17:00 (Not established)

Thursday 28

Thu.A1 : Remote sensing II	Thu.B1 : Decision support systems I
Thu.A1 : 09:10 – 09:30 A tree canopy counting method for precision forestry *Juan Pablo Garella, Matias Tailanián, Gabriel Lema, Javier Regusci, Germán Fernandez Flores, Mónica Almansa, Pedro Mastrangelo and Pablo Musé*	Thu.B1 : 09:10 – 09:30 SIMAGRI: An Agro-climate Decision Support Tool *Eunjin Han, Walter E. Baethgen, Julieta Souza, Mercedes Berterretche Adaime, Gonzalo Antúnez, Carmen Barreira and Flora Mer*
Thu.A1 : 09:30 – 09:50 SOC IoT data collection platform: application to oceanic temperature sensing *Ariel Sabiguero and Angel Segura*	Thu.B1 : 09:30 – 09:50 A Decision Support System for Fish Farming using Particle Swarm Optimization *Angel Cobo, Ignacio Llorente, Ladislao Luna and Manuel Luna*
Thu.A1 : 09:50 – 10:10 Design of a low power wireless sensor network platform for monitoring in citrus production *Leonardo Steinfeld, Javier Schandy, Federico Favaro, Andrés Alcarraz, Juan Pablo Oliver and Fernando Silveira*	Thu.B1 : 09:50 – 10:10 Implementation of Robust Decision Making in Agriculture Planning Decisions using Cloud Computing *Xavier Gonzalez*
Thu.A1 : 10:10 – 10:30 Development of a wireless sensor network system for the monitoring of insect pests in fruit crops *Leonardo Barboni, Fernando Silveira and Alvaro Gomez*	Thu.B1 : 10:10 – 10:30 Decision support system for farmland fertilization based on linear optimization with fuzzy cost *Esmelin Niquín Alayo, Edmundo Vergara Moreno and Marks Calderón Niquin*

Thu.A2 : Optimization and simulation III	Thu.B2 : Artificial intelligence III
Thu.A2 : 14:30 – 14:50 Assessing traceability system adopted by the Mango supply chain in Colombia: An analysis of the asynchrony in the inventory and food quality *Milton Herrera and Javier Orjuela*	Thu.B2 : 14:30 – 14:50 Forecasting Pesticides Usage Trends Based on Evolutionary Scientific Workflows *Sergio Serra, Anderson Oliveira, Fabricio Farias and Raimundo José Macário Costa*
Thu.A2 : 14:50 – 15:10 A Simulation Model to Analyze the Payback Period of a Sow Farm Using the Transient State *Marco Antonio Montufar and David Fernando Munoz*	Thu.B2 : 14:50 – 15:10 Spatial variability inside a greenhouse can be modeled with machine learning *Vinicius A. V. Lopes, Felipe F. Bocca and Luiz Henrique A. Rodrigues*
Thu.A2 : 15:10 – 15:30 Simulation of cattle farms with System Dynamics in a serious videogame. Case: SAMI *Urbano Eliécer Gomez Prada, Oscar Gómez*	Thu.B2 : 15:10 – 15:30 Neglecting autocorrelation in development degrades performance of sugarcane yield models *Matheus A. Ferraciolli, Felipe F. Bocca and Luiz Henrique A. Rodrigues*
Thu.A3 : Data envelopment and multicriteria	**Thu.B3 : Data mining and bussiness intelligence II**
Thu.A3 : 16:00 – 16:20 A multiobjective model to determine the sustainability level of livestock production in the Huascaran National Park *Jesús E. Espinola, Henry A. Garrido, Angel Cobo, Fernando Salmón, Edwin J. Palomino, Esmelin Niquin and Maximiliano E. Asís*	Thu.B3 : 16:00 – 16:20 Review of Data mining applications in forestry sector *Diego Broz, Alejandro Olivera, Victor Viana Céspedes and Daniel Rossit*
Thu.A3 : 16:20 – 16:40 Using a multiobjective DEA model to assess the eco-efficiency of organic blueberry orchards in the CF+DEA approach *Lidia Angulo Meza, João Carlos Soares de Mello, Alfredo Iriarte, Marcela González-Araya and Ricardo Rebolledo-Leiva*	Thu.B3 : 16:20 – 16:40 Application of data mining to forest operations planning *Daniel Rossit, Alejandro Olivera, Víctor Viana Céspedes and Diego Broz*
Thu.A3 : 16:40 – 17:00 Using CF+DEA method for assessing eco-efficiency of Chilean vineyards *Ricardo Rebolledo-Leiva, Carlos Rodríguez-Lucero, Melany Campos-Rojas, Eduardo Pacheco-Rojas, Marcela González-Araya, Alfredo Iriarte and Lidia Angulo Meza*	Thu.B3 : 16:40 – 17:00 Business Intelligence technologies for the automation and analysis of meteorological parameters for agriculture in Ancash-Peru *Rocío Rocha, Jesús E. Espinola, Angel Cobo, Rafael Figueroa, Lluis M. Plá and Maximiliano E. Asís*

Friday 29

Fri.A1 : Descriptive mathematical models	Fri.B1 : Technological developments
Fri.A1 : 09:00 – 09:20 Time-dependent performance evaluation of a tire repair system in the agricultural stage of sugarcane industry *Carolina Gualberto, Lasara Rodrigues and Reinaldo Morabito*	We.B1 : 09:00 – 09:20 SIGRAS App: climate, vegetation and soil information for support systems for decision making in agricultural production through smart devices *Guadalupe Tiscornia, Agustín Gimenez, Adrián Cal and José Pedro Castaño*
Fri.A1 : 09:20 – 09:40 Game theory concepts and changes in the Brazilian agriculture *Fernando L. Garagorry*	Fri.B1 : 09:20 – 09:40 Number, maps and facts: Agriculture leads environmental preservation *Evaristo Eduardo De Miranda, Carlos Alberto De Carvalho, Osvaldo Tadatomo Oshiro, Paulo Roberto Rodrigues Martinho, Lucíola Alves Magalhães and Gustavo Spadotti Amaral Castro*
Fri.A1 : 09:40 – 10:00 A Stochastic Frontier Approach in the Presence of Endogeneity for the Brazilian Agriculture *Geraldo Souza and Eliane Gomes*	Fri.B1 : 09:40 – 10:00 Generating spatial data of Brazilian social vulnerability *Luciola Alves Magalhães, Marcelo Fernando Fonseca, Davi de Oliveira Custódio, Paulo Roberto Rodrigues Martinho, Jaudete Daltio, Carlos Alberto de Carvalho and Gustavo Spadotti Amaral Castro*
Fri.A1 : 10:00 – 10:20 Geostatistical study of root rot produced by the fungus Rhizoctonia solani Kühn in the cultivation of Vigna unguiculata (L.) Walp in municipality of Gibara, Cuba *Vilma López Cruz, Esteban López Milán, José Quintín Cuador Gil and Ramón Candelario Núñez Tablada*	Fri.B1 : 10:00 – 10:20 IntegraGIS: A GIS system that integrates habitat modelling for vegetable species and the 3PG growth predicting model *Adrián Márques, Marcelo Ortelli, Alvaro Pardo, Francisco Rodríguez, Diego Strasser, Pablo Hernández, María León, Marcela Rodriguez and Daniel Valdomir*
Fri.A1 : 10:20 – 10:40 Sugarcane Yield Estimate Analysis by using Regression Error Characteristic Curves (REC Curves) *Luiz Henrique A. Rodrigues and Felipe F. Bocca*	Fri.B1 : 10:20 – 10:40 A GIS system to prevent country-wide soil erosion and support sustainable agriculture *Walter Díaz, Adrián Márques, Alvaro Pardo, Javier Preciozzi, Santiago Arana and Gervasio Piñeiro*

Fri.A2 : Optimization and simulation IV	Fri.B2 : Decision support systems II
Fri.A2 : 11:10 – 11:30 Production Planning Model for the assignment of Fermentation Tanks at Wineries *Carlos Monardes, Alejandro Mac Cawley, Jorge Vera, Susan Cholette and Sergio Maturana*	Fri.B2 : 11:10 – 11:30 A Decisions Support System for Purchasing and Storing Fresh Fruit *Wladimir Soto-Silva, Marcela González-Araya and Lluis Pla-Aragonés*
Fri.A2 : 11:30 – 11:50 Integrated model of crop rotation planning and delineation of rectangular management zones *Virna Ortiz-Araya, Víctor M. Albornoz and Rodrigo A. Ortega*	Fri.B2 : 11:30 – 11:50 Research Directions in Technology Development to Support Real-Time Decisions of Fresh Produce Logistics *J. René Villalobos, Wladimir Soto-Silva, Marcela González-Araya and Rosa Guadalupe Gonzalez Ramirez*
Fri.A2 : 11:50 – 12:10 Resolution of Mixed-Integer Bilevel Problem in the supply chain in meat industry by an Branch & Bound Algorithm *Camila Flores, Victor Albornoz, Sara Rodríguez and Manuel Jiménez-Lizágarra*	Fri.B2 : 11:50 – 12:10 A new cloud decision support system for tactical planning in a fruit supply chain *Jordi Mateo Fornes, Wladimir Soto, Marcela Gonzalez, Adela Pagès Bernaus, Lluís Miquel Plà Aragonés and Francesc Solsona*

“Hay Campo para las TICs” - Jornadas Uruguayas de Nuevas Tecnologías para el Agro

ICT4V Information and Communication Technologies for Verticals

Agenda

UNIVERSIDAD DE LA REPUBLICA URUGUAY — FACULTAD DE INGENIERIA

	Miércoles 27 de setiembre	Jueves 28 de setiembre	Viernes 29 de setiembre
14:00	Apertura oficial del evento **Presentación invitada: "Perspectiva de Internet de las Cosas en un ambiente rural"** *Thiago Santos - Embrapa (Brasil)*	**Presentación invitada: "Innovaciones en AgroTICs en América Latina"** *Ana Castillo - BID-FOMIN*	**Presentación invitada: “Plataforma Agro Big Data”** *Dr. Esteban Feuerstein, Director Fundación Sadosky (Argentina)*
	"RFID e IOT en Agroindustria" *Bruno Bellini - BQN*	"Perspectivas de las AgroTICs en Uruguay: integrando el Estado y el sector productivo" *Mercedes Berterreche*	"Los sistemas de información geográfica aplicados al agro" *Pablo Rebufello, Pablo Piperno – ICA*
	"Soluciones de IoT para el campo basadas en LPWAN" *Germán Capdehourat - Teliot*	"Sistemas nacionales de información y genómica en la mejora de calidad del canal" *Elly Navajas - INIA*	"Nuevas Tecnologías para la gestión de la viticultura uruguaya: experiencia y aplicación" *Milka Ferrer - Facultad de Agronomía, UDELAR*
15:30	Break	Break	Break
16:00	"Monitoreo y diagnóstico del entore: un ejemplo de ganadería extensiva de precisión aplicada a la reproducción bovina" *J. Oreggioni - Agromote*	"Incorporación de RFID para gestión de procesos en fase primaria lechería" *COLAVECO*	"Decisor CREA: Herramientas para la sostenibilidad ambiental y económica en el agro uruguayo" *Juan Ignacio Buffa – FUCREA-FOMIN*
	"El tambo robótico como etapa evolutiva de las tecnologías de ordeñe en Uruguay" *Ignacio Torres- GEA*	"Aplicaciones de productos fitosanitarios: Nuevas Tecnologías para el uso responsable y cuidado del medio ambiente" *Gastón Lieutier - Fumigapp*	"Herramientas cognitivas como apoyo para la toma de decisiones en la producción agropecuaria" *Gerardo Alvarez, Fernando Lopez Bello- Quanam*
17:00	**Keynote: "Agricultura digital: de la biotecnología al Big Data / Agricultura inteligente y sustentable"** *Carlos Meira - Embrapa (Brasil)*	**Keynote: "Las TICs en diferentes actividades del sector agropecuario: investigación, políticas públicas y producción"** *Walter Baethgen – Senior Scientist, International Research Institute for climate and society, Columbia University, USA.*	**Panel de cierre: "Innovación en el Agro: aporte de las TICs, presente y futuro"** *María Simón (Decana de la Facultad de Ingeniería), Santiago Dogliotti (director ANII), Agustín Giménez (director GRAS INIA), María Eugenia Oholeguy (Presidente FUCREA), Eduardo Barre (Director de la Dirección General Servicios Ganaderos MGAP)*

LUGAR: Anfiteatro del Edificio Polifuncional José Luis Massera, anexo de la Facultad de Ingeniería. Av. Julio Herrera y Reissig 565 – Montevideo- Uruguay

Proceedings of the

1st International Conference on Agro Big Data and Decision Support Systems in Agriculture

Editors:

Héctor Cancela, Universidad de la República, Uruguay
Lluís Miquel Plà-Aragonés, Universitat de Lleida, Spain
Antonio Mauttone, Universidad de la República, Uruguay
Adela Pagès-Bernaus, Universitat de Lleida, Spain

Digital Edition: Universidad de la República - Universitat de Lleida
September 2017
ISBN: 978-9974-0-1514-2

Table of Contents

Remote monitoring of physico-chemical water parameters in rural water sources

J. Mirón[1], M. Grande[1], M. Pohl[1], F. Huguet[1]

[1] Universidad Centroamericana "José Simeón Cañas", UCA
Bulevar Los Próceres, San Salvador, El Salvador
fhuguet@uca.edu.sv

Abstract

We present the description of a telemetry system for monitoring physico-chemical water variables using the *General Radio Service* (GPRS) mobile communication technology for data transmission. The system is intended to monitor rural water sources feeding communities without potable water service. The telemetry system is linked to a web service for consultation and analysis of the recorded data, and to a geographic information system for water resources management.

1. Introduction

In El Salvador, the institution in charge of the network of aqueducts and sewers is the *Administración Nacional de Acuedúctos y Alcantarillados*, ANDA (National Administration of Aqueducts and Sewers). Despite efforts to cover most of the country's inhabited regions with an aqueduct system and potable water service, this has not yet been achieved, especially in rural areas. For this reason, in areas without potable water distribution services, the population self-supplies directly from natural sources (rivers and wellsprings) and rainwater collectors. In almost all of these cases, there is no monitoring of water parameters indicating its suitability for consumption, nor a history of data of the hydric resource behavior allowing its study in order to establish treatment and management plans.

One of the possible solutions to this problem is the implementation of monitoring systems composed of remote units scattered throughout areas having rural water sources, and a data collecting centralized unit. This paper concerns the design and implementation of a telemetric monitoring system of physico-chemical variables of water for rural sources (Mirón, Grande & Huguet, 2017). It has been developed by a joint initiative of the local government of the municipality of Tecoluca, located in a central rural area of the country, and the *Department of Electronics and Computer Science of the Central American University "José Simeón Cañas"*. In the medium term, it is planned to gradually install the remote monitoring units in the water sources of the town.

2. System description

Searching for the best compromise between cost and performance, the monitoring system incorporates Arduino technology, the use of an autonomous photovoltaic power system, the use of the public telephone network for data transmission, and the use of free software for the reception, analysis, manipulation and data storage. The general system is composed of two parts:

- *Remote monitoring unit.*
- *Centralized data collector unit.*

The remote monitoring unit is composed of 3 modules:

- *Control process and data acquisition module.* This module is the core of the remote monitoring device. It controls the data acquisition, storage and the transmission process. It incorporates an Arduino Uno board, a SD memory shield, an external clock, and a set of sensors for measuring flow, temperature, salinity, pH and dissolved oxygen. In further implementations turbidity and chlorophyll sensors will be included. The Arduino board contains the main program that executes temporized sensors lectures, data storage and transmission tasks. Also, the main program has a user interface routine for sensors calibration and for setting operation parameters.

- *Data transmission module*. It uses the General Packet Radio Service technology (GPRS) for sending data through the TCP/IP protocol (Transmission Control Protocol/Internet Protocol) to the receiver server connected to the public network. It incorporates a FONA 800 shield.
- *Photovoltaic power supply module*. In addition to the constraints concerning the distribution of drinking water in rural communities, it is common to find a limited electric power distribution service. For this reason, the design of the remote monitoring unit includes a photovoltaic power module to make it self-sustaining. Figure 1 shows the implemented remote monitoring unit.

Figure 1: Implemented remote monitoring unit

The centralized data collector unit consists of a server connected to the web hosting an ensemble of scripts for reception, storing and data display. The server receives the data through an HTTP-GET packet (a Hypertext Transfer Protocol request method), and unpacks and processes it using the Hypertext Pre-processor scripting language (PHP). The data is stored in a Structured Query Language (SQL) database. The database incorporates a query and report generation system and is linked to a previously implemented geographic information system (Zepeda, Deras, Juárez & Quintanilla, 2016).

3. Data flow

The data flow consists of different stages through which the collected information passes before being displayed. These steps are the following:

- *Data capture, storage and package*. Signals arriving at the input ports of the Arduino board are read, digitalized and converted into structured text strings with annexed information concerning the reading time code and the IP address of the receiving server. This strings are also stored inside the microSD memory.
- *Data transmission*. The FONA device initiates the GPRS communication. Once established, the *Data transmission module* generates calls to the receiver server IP using the GET method and the HTTP protocol. The previously structured data string is concatenated in the URL (Uniform Resource Locator) as a send variable of the HTTP call. After sending the data, the *Data transmission module* keeps waiting for a reception confirmation message from the server.
- *Data reception*. Once the HTTP-GET packet is received by the server, it is unpacked and processed using PHP. The encoded string is extracted and a new record is inserted into the database. If the reception process is successfully completed, the server sends a value of 1 to the *Data transmission module*. Otherwise, a value of 0 is sent.
- *Confirmation of received data*. The remote unit receives the confirmation message sent by the server and, depending on the value, stores the data in the microSD card. This record is a register

of the amount of successful sendings. Recorded data strings concerning failed transmissions are labeled inside the microSD memory.

4. Results

The following sections of the paper focus on the data transmission process.

Field tests were done at one of Tecoluca's water sources in order to determine the data transmission performance. Because of the weak signal of telephone companies providing GSM/GPRS services in this rural zone, the percentage of successful sending in the first test was only 75 %. To enhance this performance, the transmission routine of the remote monitoring unit was modified. The number of GPRS search attempts was raised from 7 to 15. Also, the number of sending package attempts was raised to 3. With these changes, the transmission task performance raised to 88% on site.

5. Conclusions

The use of the public mobile network for telemetry systems is economically profitable. The maximum savings are achieved in the low-cost infrastructure needed to use this technology. Similarly, in recent years the cost of network usage contracts has decreased making this option more feasible. A disadvantage in this method is the risk of failures or signal limitations of the contracted network, however, the GPRS network systems and their upgrades are increasingly stable and have enhanced coverage. To assure reliable operation of telemetry systems in rural areas using the public mobile network, a verification of the best provider option in terms of mobile signal level at installation places must be performed. A previous consultation with the contractor of the telephone company is recommended to know about network expansion projects in the future.

In addition to the considerations regarding technical communication problems, it is necessary to take into account other complexities linked to telemetry systems. These complexities concern the protection and maintenance of the system, and the interpretation and management of the collected data. Concerning protection, it is necessary to envisage an infrastructure that protects the remote monitoring units from vandalism, such as theft of elements or damage. Regarding maintenance, it is necessary to contemplate a periodic investment in replacement parts and in technical personnel trained for revision and service. At the same time, it is necessary to count with water specialists or search for collaborations with institutions related to the subject to ensure the correct interpretation and exploitation of the collected information.

References

[1] Mirón J, Grande M, Huguet F. (2017). *Dispositivo electrónico para la captación de variables físico químicas del agua* (Unpublished degree thesis). El Salvador: Universidad Centroamericana "José Simeón Cañas" UCA.

[2] Zepeda C, Deras E, Juárez C, Quintanilla L. (2016). *Sistema de monitorización para la calidad de agua en el municipio de Tecoluca* (Unpublished degree thesis). El Salvador: Universidad Centroamericana "José Simeón Cañas" UCA.

Remote Sensing of Algal Blooms in the Uruguay River Based on Multispectral Satellite Imaging and Field Data

José Lezama[1*], Fernanda Maciel[2], Francisco Pedocchi[2], Pablo Musé[1]

[1] IIE, Universidad de la República, Uruguay

[2] IMFIA, Universidad de la República, Uruguay
* jlezama@fing.edu.uy

Abstract

Algal blooms in freshwater bodies can have a major negative impact in humans and the aquatic life in general. Remote sensing of water eutrophication using multispectral satellite imagery is a powerful tool for analyzing its causes. Using public algal bloom records measured along the Uruguay River and Landsat-8 multispectral images, we learn a new algal concentration index. We demonstrate that traditional vegetation indexes such as NDVI, FAI and others do not correlate well with algal blooms, and conclude that it is better to learn from local data. The result is a refined tool for analyzing the eutrophication in the Uruguay River.

1 Introduction

Blooms of certain algae, such as cyanobacteria, can generate toxins that are potentially harmful for both humans and aquatic life. According to [1], Salto Grande reservoir in the Uruguay River is one of the areas with the greatest risk of exposure to cyanobacterial blooms in the country. They are often associated with nutrient enrichment and hydrological changes ([1]). Additionally, blooms occur more frequently in summer, with higher temperatures and low flow conditions ([3]). Better understanding the causes and prediction of algal blooms are both important challenges for environmental studies.

Satellite imagery provides a cost-effective way to monitor large water bodies. For instance, crossing information of satellite imagery and agricultural or industrial activities in the basin, could help identify sources of excessive nutrients conveyed into the water body.

This study focuses upon the lower Uruguay River, located between Argentina and Uruguay, from Salto Grande reservoir to the river mouth in Río de la Plata estuary. Using a publicly available record of in-situ algal bloom measurements taken on 35 CARU[1] monitoring stations all along the Uruguay River (Fig. 2), and publicly available Landsat-8 satellite imagery, we learn a new index to estimate occurrence of algal blooms in the Uruguay River.

2 Data Acquisition

2.1 Satellite Imagery

Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images ranging from January 2014 to April 2017 were used for this study[2]. We used the top of the atmosphere (ToA) reflectance values for bands 1 to 7 (blue to short wave infrared, 430 to 1650 nm) and also include the thermal bands 10 and 11 (10,600 to 12,510 nm). These bands have a resolution of 30 meters.

We use satellite images when the date of satellite overpass is no more than one day before or after the date the in-situ measurements were performed. We also discard satellite scenes with significant cloud cover. We obtain a total of 13 usable scenes that satisfy this criterion and can be used for training and validation, and a total of 34 scenes for evaluation.

[1]CARU stands for “Comisión Administradora del Río Uruguay”

[2]Downloaded from the United States Geological Survey (USGS) website

predictor name	coefficient	p-value	t-value	SE
Intercept	222.8645	1.55e-01	1.47	150.90
Sun elevation	33.4551	1.64e-02	2.56	13.07
Band 1 (Coastal Aerosol)	-32.2250	1.05e-03	-3.44	9.40
Band 2 (Blue)	624.9488	3.75e-01	0.93	673.08
Band 3 (Green)	-929.9672	3.25e-01	-1.03	903.20
Band 4 (Red)	-606.4420	2.67e-02	-2.30	263.60
Band 5 (NIR)	355.6100	1.97e-05	4.43	80.39
Band 6 (SWIR 1)	-77.6375	4.80e-01	-0.74	104.42
Band 10 (TIRS 1)	1.5566	9.43e-04	3.46	0.45
Band 11 (TIRS 2)	24.2003	6.95e-03	2.87	8.48

Table 1: Statistical significance of the predictor variables. Coefficients are not standardized.

2.2 In-situ Measurements

The in-situ algal bloom measurements are obtained from public reports from CARU, downloaded from CARU's website[3]. The reports are in PDF format, with dated and georeferenced measurements of algal concentration for 35 different monitoring stations. The algal concentration is reported in three categories: "No algal bloom", "Use caution" and "Do not bath". Given the limited amount of available data, in order to simplify the classification problem, we binarized the labels by grouping the latter two categories into one, obtaining one negative and one positive class.

Using a custom PDF scraping script, we extracted from these reports a total of 1422 datapoints. We restricted them to those that are within one day of the satellite overpass, finally obtaining 100 negative and 22 positive datapoints to train and validate our index.

3 Model

We used a logistic regression on the aforementioned satellite bands to learn the negative and positive categories from CARU reports. The logistic regression has three main advantages in this setting. First, its underlying linear nature makes overfitting less likely, an important consideration given that training samples are scarce in our case. Second, it allows some interpretability on the importance of each of the spectral bands in the estimator. Third, it allows a computationally efficient evaluation of the model in large images of several megapixels.

Because the algal blooms are known to be seasonal events in the river ([3]), we also use as a predictor variable the sun elevation angle during the satellite overpass which, because Landsat-8 orbit is sun-synchronous, encodes the time of the year.

We perform 10-fold cross-validation to validate the model on data not seen during training. Figure 1 presents the ROC curves for training and validation. We compare with the classic NDVI index [6], the Floating Algae Index (FAI) [5], with a simple green/red index [4] and with a previous study in the Uruguay River [2]. In Table 1 we present the statistical significance of each band in the estimator.

4 Experimental Results

Figure 2 shows the obtained index overlaid on RGB satellite images of the Uruguay River for different dates. One can clearly observe that blooms tend to occur during the austral summer and that they are stronger in quiet waters, such as the reservoir of the Salto Grande dam (near the top of the figure).

5 Conclusions

Compared to indices proposed elsewhere, we obtained a better algal bloom estimation index derived from satellite imagery, with respect to in-situ measurements for the Uruguay River. The proposed index

[3] `http://www.caru.org.uy/web/informesalgas/`

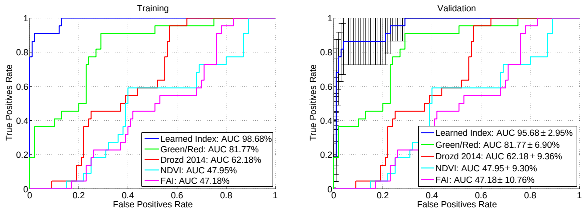


Figure 1: ROC curves for algal bloom detection for training (left) and validation (right). Validation ROC curves are averaged over 10 cross-validation folds, showing the model's robustness to overfitting.

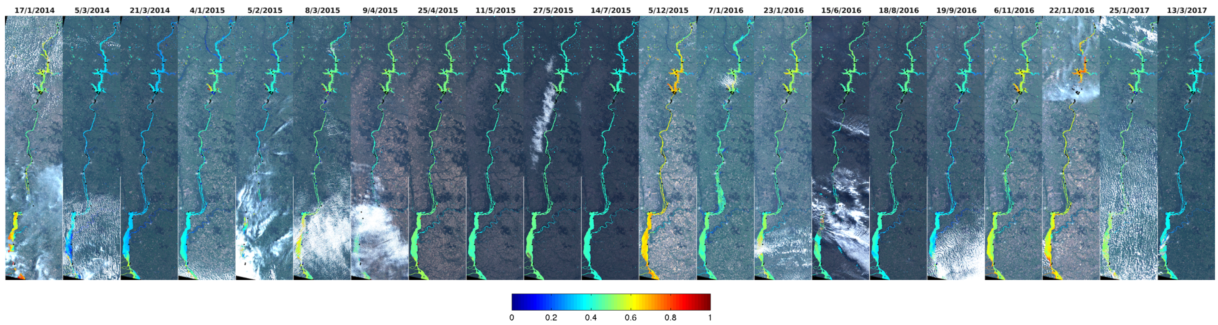


Figure 2: Evaluation of our proposed index for 21 clean scenes of the Uruguay River ranging from January 2014 to March 2017. Date for each scene is shown on top. The index can be interpreted as the probability of the presence of algae in the water. Black dots represent the measuring stations. Blue and red dots represent the locations in space and time of the negative and positive samples respectively, that were used for training. Best viewed in electronic format.

could be used to identify areas in the river that are more prone to algal blooms, such as water reservoirs, and study the historical causes of eutrophication. We observed that the thermal bands of Landsat-8 are useful for algal bloom estimation, probably because they encode the water temperature, which is one of the key variables associated to cyanobacterial blooms ([3]). We expect the index to improve when more in-situ measurements aligned with satellite overpass are available as additional training data.

References

[1] S. Bonilla, S. Haakonsson, A. Somma, A. Gravier, A. Britos, L. Vidal, L. De León, B. M. Brena, M. Pírez, C. Piccini, G. Martínez de la Escalera, G. Chalar, M. González-Piana, F. Martigani, and L. Aubriot. Cyanobacteria and cyanotoxins in freshwaters of Uruguay. *INNOTEC*, 10:9–22, 2015.

[2] A. A. Drozd, G. Ibáñez, F. Bordet, and S. E. Torrusio. Remote sensing estimation of chlorophyll "a" concentrations on hypertrophic waters using hyperspectral field spectrometer, SPOT-4 (HRVIR) and Landsat-7 ETM+ data. 2014.

[3] G. Ferrari, M. d. C. Pérez, M. Dabezies, D. Míguez, and C. Saizar. Planktic cyanobacteria in the lower Uruguay River, South America. *Fottea*, 11(1):225–234, 2011.

[4] H. R. Gordon and A. Y. Morel. *Remote assessment of ocean color for interpretation of satellite visible imagery: A review*, volume 4. Springer Science & Business Media, 2012.

[5] C. Hu. A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, 113(10):2118–2129, 2009.

[6] J. Rouse Jr, R. Haas, J. Schell, and D. Deering. Monitoring vegetation systems in the great plains with erts. 1974.

Sensor Data Analysis and Sensor Management for Crop Monitoring

Raquel Sosa[1], Andres Vera[1], Maite Ibarburu[1], Leonardo Steinfeld[2], Javier Schandy[2], Fernando Silveira[2]

[1] Universidad de la República
Julio Herrera y Reissig 565 - Instituto de Computación
raquels@fing.edu.uy

[2] Universidad de la República
Julio Herrera y Reissig 565 - Instituto de Ingeniería Eléctrica
{leo,jschandy,silveira}@fing.edu.uy

Abstract

During crop growth there are microclimate variations that can affect them. Some variations, such as slight frost, if repeated many times can affect the yield of the crops. These variations can be monitored by a wireless sensor network and the estimated yield of the crop can be adjusted by analyzing the data. This work proposes a system that manages a wireless sensor network to monitor environmental variables, stores the data it receives and enables different types of users to analyze data. The users can be network admins or agronomists. The proposal includes the use of Geographic Information Technologies to display the data.

1 Introduction

During the crop growth microclimate variations could occur in different zones of the fields and affect the yield of the crop. Agronomists could learn about the real effect of those variations on crops if they had detailed data about them. To really know about those variations the deployment of a wireless sensor network (WSN) on the field can take measurements of environment variables, like temperature or humidity. The development of that kind of WSN applied to agriculture is part of a research project of the Electrical Engineering Institute [1] (IIE)[3] and is supported by INIA[2].

In the last few years there has been a collaboration between the IIE and the Computer Science Institute [3] (InCo) on different uses of the data provided by the WSN on the field. The main user of the data generated by the WSN is the agronomist who needs to analyze them. Besides, the network administrator needs to monitor the WSN status and know how it is working through time during the test phase of the deployment.

There is some work on the analysis of sensor data, like [2] that focuses on the scientific applications of environmental data provided by several data sources, including meteorological stations. Its applications are terrestrial ecology and oceanography. This work is based on the IIE research project considering agronomists as one of the main users of the system proposed but also considering the needs of the researcher on networks.

2 Wireless Sensor Network

Wireless sensor networks [4](WSN) comprised sensor nodes that measure the environment and send the information to a root node. A key feature is the ad-hoc formation of a mesh network, where all nodes can route information to the root node.

The WSN was developed based on free and open-source software (FOSS). The protocols adopted for the communication stack are standardized by IEEE and IETF. The sensor nodes embedded software was

[1]Instituto de Ingeniería Eléctrica - Facultad de Ingeniería - http://iie.fing.edu.uy/
[2]Instituto Nacional de Investigaciones Agropecuarias - http://www.inia.uy
[3]Instituto de Computación - Facultad de Ingeniería - http://www.fing.edu.uy/inco

build using Contiki [4], an event-driven operating system oriented to WSN and Internet of Things (IoT) applications using constrained hardware.

The communication stack adopted is the full-stack usually known as 6LoWPAN, since it uses IPv6 over IEEE 802.15.4 wireless personal network (low power WPAN).

The physical and MAC layer is based on the IEEE 802.15.4 standard operating in the 2.4 GHz unlicensed band. The access mechanism is Carrier Sense Multiple Access with Collision Avoidance (CSMACA) where the node does carrier sensing before transmitting packets to check whether the channel is idle or not. The upper layers are standardized by the IETF. In order to transport of IPv6 packets over 802.15.4 links, the 6LoWPAN adaptation layer protocol is used.

On top of 6LoWPAN, RPL is adopted (IPv6 Routing Protocol for Low-Power and Lossy Networks), a proactive routing protocol based on a tree-oriented strategy. The distance metric is usually based on some link quality indicator. RPL enables different operation modes. In this case, each network node's storing the default route to the root and table entries to route packets to all the nodes downwards the tree was selected.

At the application layer the Constrained Application Protocol (CoAP) [5] is used, which is a RESTful protocol for use with constrained hardware such as WSN nodes. It relies on UDP on transport layer. CoAP follows a REST model in which the nodes, as servers, make resources available to clients under a URL. In this case the client is the root node which is connected to a gateway that sends the information to a server via cellular network. CoAP uses methods such as GET, PUT, POST, etc. and a special OBSERVE mechanism that allows client nodes to retrieve a resource value from a server without a explicit GET.

The WSN generate data from their sensors at variable (configurable) rates and these date are transmitted to the base node through different routes across several nodes. The base node has limited storage capacity and limited computing power, but it is the node that connects the network to the internet. The base node receives all the data collected by the node sensors including the battery level and routing information.

3 System Description

This work focuses on the system that communicates with the base node and manages the data generated by the WSN. The system proposed also manages the data on the status of the WSN and the nodes, allowing the WSN administrators to monitorize the network and set parameters. Some of the functionalities required of the system were the georeferencing of the node position and the spatial analysis of the data provided, mainly of the WSN status data and in relation with other spatial data (hills, roads, wind map). We propose the graphical analysis of the variables measured (air humidity, soil humidity, temperature) and also the spatial analysis. One interesting function of the system is the capacity of setting alarms associated with threshold values for the observed variables (for example soil humidity below 30% could mean a drought and the agronomist could reinforce artificial irrigation). Some non-functional considerations of the system are that it has to be connection fault tolerant, considering that the base node has low computing power and has to store historical data (several years).

To achieve the needs of data analysis and WSN management, given that this kind of network is in an development phase, we designed a system with a distributed architecture and using several technologies that enable the extension of the system. As shown in Figure1, the system has a component (sensors-daemon) that receives the data from the network. The sensors-daemon runs in a limited hardware component near the network, provides temporal storage of the data, and communicates the sensors-core with the network. Sensors-core and sensors-daemon have two ways of communicating.One of them is asyncronous to deal with internet connection losses and efficient use of the bandwidth. When the system needs to handle serveral WSN in different deployments, the component sensors-daemon will be replicated near each network. Sensors-core runs in a server and stores the data using two types of databases:

[4] www.contiki-os.org

a relational one for system operation and a non-relational one for storing measured data. This second database allows the system to use diverse analysis tools related to Big Data technologies.

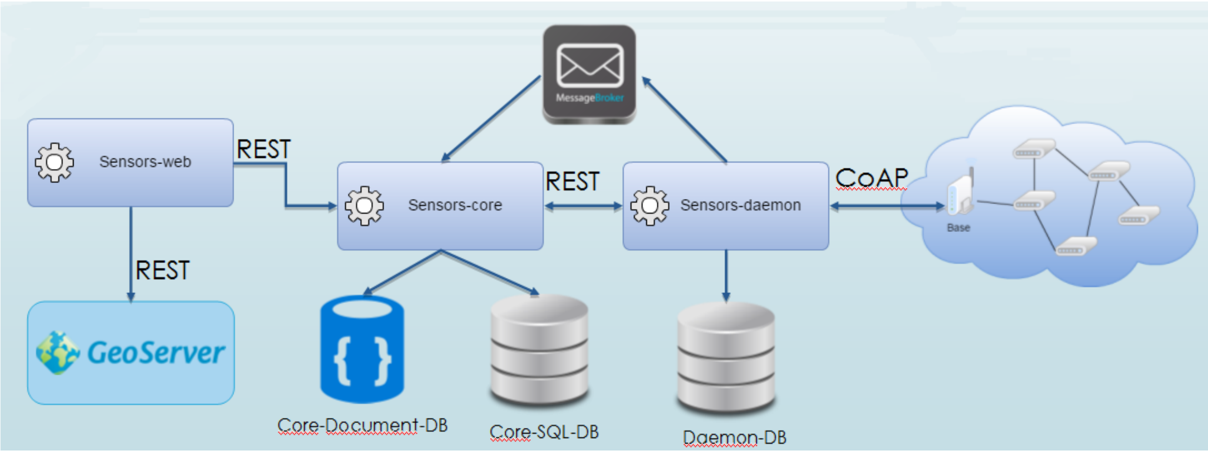


Figure 1: System Arquitecture

The system provides the different users with a web interface so that they can access the system through the Internet. The sensor-web component uses a MapServer (GeoServer[5]) to overlay sensor data with context spatial data in a flexible way. Geoserver also provides the standar Web Processing Service (WPS)[1] that allows the system a further spatial analysis of the data.

4 Conclusions and Future Work

This work is the first prototype of a system that allows users to analyze crop field data in terms of time and location. From the very beginning of the project, the possibility of generating alarms for users when unusual situations are detected was considered (for example, very high temperatures which could be indicators of fire). Our proposal relates the data gathered by the sensors with its spatial location, allowing users to perform different types of GIS analyses (such as heatmaps) and cross it with other spatial data. The inmediate future work is to test the system with real test fields since the system reached a alfa test in the IIE.

References

[1] Ogc wps 2.0 interface standard, 2015.

[2] Derik Barseghian, Ilkay Altintas, Matthew B. Jones, Daniel Crawl, Nathan Potter, James Gallagher, Peter Cornillon, Mark Schildhauer, Elizabeth T. Borer, Eric W. Seabloom, and Parviez R. Hosseini. Workflows and extensions to the kepler scientific workflow system to support environmental sensor data access and analysis. *Ecological Informatics*, 5(1):42–50, 2010.

[3] Fernando Silveira, Leonardo Barboni, Leonardo Steinfeld, Pablo Mazzara, and Alvaro Gómez. Gervasio : Generalización de las redes de sensores inalámbricos como herramienta de valorización en sistemas vegetales intensivos, jun 2017.

[4] Shuang-Hua Yang. *Wireless Sensor Networks: Principles, Design and Applications*. Springer London, London, 2014.

[5] Carsten Bormann Zach Shelby, Angelo P. Castellani. Coap: An application protocol for billions of tiny internet nodes. *IEEE Internet Computing*, vol. 16:pp. 62–67, 2012.

[5] GeoServer - http://geoserver.org

Large data volume parallel clustering for computational cost reduction in Self-Organized Networks training

Mariela Azul González[1], Pablo Montini[2], Jorge Martínez Arca[1], Lucía Isabel Passoni[1]

[1]Laboratorio de Bioingeniería. ICyTE. Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (CONICET: Consejo Nacional de Investigaciones Científicas y Técnicas).
Universidad Nacional de Mar del Plata
Juan B. Justo 4302, Mar del Plata, Buenos Aires
mazulgonzaleza@fi.mdp.edu.ar, amartinez@fi.mdp.edu.ar,lpassoni@fi.mdp.edu.ar

[2] *Grupo de investigación de Inteligencia Artificial aplicada a Ingeniería*
Facultad de Ingeniería
Universidad Nacional de Mar del Plata
Juan B. Justo 4302, Mar del Plata, Buenos Aires
montini@fi.mdp.edu.ar

Abstract

Reducing the computational cost is of great importance in applications that handle large volumes of data (Data Mining, Big Data, etc.). In cases of large data volumes, the computational cost of Artificial Neural Networks training increases, limiting the efficiency of the whole process. The aim of this work is to compare different process pipelines oriented to segment regions of interest in video sequences of biological dynamic patterns. As a hypothesis, we assume the decreasing of the amount of an unsupervised neural network (Self Organizing Maps) training data allows a computational cost reduction. The computational costs are evaluated using statistical parameters of simulated experiments.

1 Introduction

Whenever Artificial Neural Networks are applied to pattern recognition tasks that deal with large data volumes, the computational cost of their training will increase, limiting the efficiency of the whole process. In this work we propose a new approach to the design of a system that segment regions of interest (ROI) in video sequences of large volume of biological dynamic patterns. As previously published, features extracted from frame sequences are used as input to an unsupervised neural network, such as self-organizing maps (SOM) to deal with recognition of ROI [1-3].

Given that in most of the agricultural applications, multiple stations are sparsely located at the collection points, the system design must consider the amount of data to be transmitted to the central processor, and it should be oriented to minimize it [4]. Several efforts had been made in order to optimize SOM techniques using large volume of data. In this direction Matharage *et al* proposed a method that trains multiple SOM networks proposed a method that trains multiple SOM networks with data partitions, and afterwards performs a projection of all SOM trained networks into a 2D grid [5]. We consider that this solution is not enough adequate because the amount of information that must be send from the local stations to the central processor is not optimized. Also, Bedregal combines the Self-Organizing Map with Metric Access Methods (MAM) during the training stage, minimizing the computational cost at the central processor [6]. This proposal improves the SOM training performance within the central station, where the total data provided by the local points must be fed, however the communication channel load between the local stations and the central one is not optimized.

In our proposal the local stations are assumed to be provided with sensors and embedded processors with capacities to acquire and process video sequences. The design of a distributed processing model will improves the computational cost of the whole pattern recognition system. Preprocessing at the local stations, where primarily data is collected, will diminish the transmission channel load. Thus SOM training with a minor data amount will reduce the whole computational complexity showing the expected effectiveness. The use of a trained SOM has shown to be a technique for coloring images developed from dynamic speckle laser videos is successfully applied to help an automatic system to identify a particular region of interest depending on the application; this presents an advantage with respect to other methods for ROI recognition [7-11]. As a particular case, we explore the goodness of the distributed design applied to ROI detection within video sequences from Dynamic Laser Speckle (DLS). DLS is observed when a surface, illuminated by a coherent light source presents some type of local activity. The intensity and shape of the observed interference pattern of scattered rays (i.e. speckles) evolve according to the sample characteristics. These speckle patterns activity is the consequence of microscopic movements or local changes in the refractive index of the sample properties. Both, the time evolution of pixel intensity as its spatial distribution over an image show seemingly random variations similar to those found in the height distributions of a rough surface. DLS patterns have been used to assess issues of interest in different fields, such as agronomy (seed analysis, fruits quality, animal sperm motility), medicine (capillary blood flow), industry (paint drying, monitoring of ice cream melting, yeast bread, gels), among others [12-15].

2 Material and Methods

In order to effectively characterize the differences of activity in the observed regions of interests (ROI) from samples using SOM, the sequence of laser intensity pattern must be preprocessed before SOM training. Several image descriptors have been developed from DLS videos in order to enhance ROI recognition. The SOM quantizes the data space of training data and simultaneously performs a topology-preserving projection of the data space onto a regular neuron (or cell) grid. SOM presents a good performance in finding natural grouping in data and was selected to process DLS sequences in this work. A Self-Organizing Map (SOM) is trained using the set of descriptors or features form DLS sequences. In order to visualize the trained SOM, first a proper coloring code for the codebook is chosen and then a Pseudo colored image is created (assigning the color of each pixel, taking the color of the SOM cell, which is the BMU for the vector of the pixel). Figure 1 briefly presents the recognition and visualization algorithm proposed.

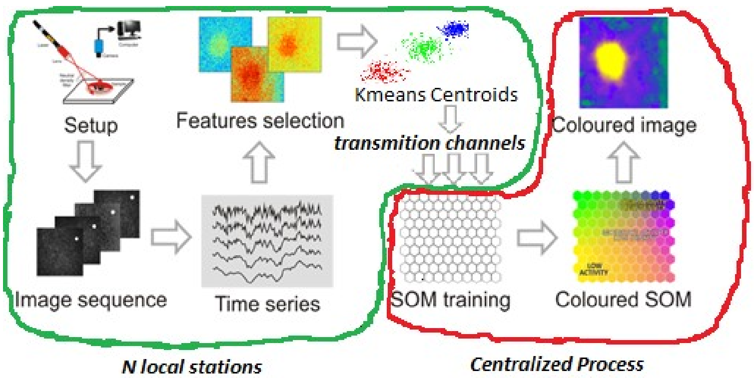


Figure 1: Proposed processing algorithm

A. Proposed Pipeline

In order to test the goodness of our proposal we compare two processes, both encompass a feature extraction stage and the training of an unsupervised neural network to be used as a recognition pattern system. A centralized process trains the SOM with the complete set of descriptors, while the proposed process trains the neural network only with the cluster centers that are generated on each distributed workstation. Hence, the optimized process (see Figure 1) presents a distributed process which comprises two steps: a first one of parallel processes that performs clustering on n descriptors sets (k-means clustering) and a second step: training a SOM with the computed k-centroids from each of the parallel working posts (n).

B. Pipeline Analysis and Evaluation

Modeling of the computational costs is performed using statistical parameters of experimental designs. Efficiency and computational cost of both pipelines applied to the segmentation of images obtained from video sequences of biological dynamic patterns from a bruise apple are analyzed. Modeling of the computational costs is performed using statistical parameters of experimental designs. Time distribution functions from the pipelines are obtained by repeating 1000 times the process of the slowest parallel work post. Consequently the simulation is designed with the bias of the worst case. GPSS World Pseudo-random number generation algorithm is used from this function to obtain computational costs statistics. That pseudo-random generator is based on Lehmer's Multiplicative Congruential algorithm, with a maximal period. The algorithm produces pseudo-random numbers in the open interval 0 to 2,147,483,647 and it generates 2,147,483,646 unique random numbers before repeating itself [16]. The Lehmer random number generator (named after D. H. Lehmer), sometimes also referred to as the Park–Miller random number generator (after Stephen K. Park and Keith W. Miller), is a type of linear congruential generator (LCG) that operates in multiplicative group of integers modulus n. The general expression is:

$$X_{k+1} = g * X_k * \bmod(n)$$

where $k:0,\dots,N$ the modulus n is a prime number or a power of a prime number, the multiplier g is an element of high multiplicative order modulus n (e.g., a primitive root modulus n), the seed X_0 is coprimeto n, and N is the total number of simulated elements. Hence, 30.000 (1000x30) simulated events were computed to broad the statistical base.

3 Results and discussion

Table 1 shows time analysis results of the proposed applied to bruise detection in an apple by DLS characterization and SOM application. The proposed method, which includes the distributed processed showed in figure 1, presented a decreasing of the amount of SOM training data and allows a noticeable time decrease when compared with the centralized process. Although it presented a computational cost reduction and could successfully detect the bruise, accuracy in borders detection slightly diminished. This fact should be detailed analyzed in cases borders accuracy is required, that was not our case of study.

	Centralized process	**Distributed process**
Mean (seg.)	2,965	8,32
Standart Desviation (seg.)	0,108	4,63
Median (seg.)	2,951	5,99

Table 1: Time analysis statistics

4 Conclusions

Modeling of the computational costs is performed using statistical parameters of experimental designs from DLS video ROI extraction rehearsals. The main contribution of this preliminary work is the reduction of the computational cost compared to the SOM training with the whole data. In future works, we will address other applications to test efficiency, computational cost requirements and data amounts.

References

[1] T. Kohonen (1995) Self-Organizing Map. Springer-Verlag.

[2] T. Kohonen, Essentials of the self-organizing map. *Neural Networks*, Volume 37, January 2013, Pages 52–65

[3] L. I. Passoni; A. L. Dai Pra; A. Scandura; G. Meschino; C. Weber; M. Guzman;H. J. Rabal; M. Trivi (2013) Improvement in the visualization of segmented areas of patterns of dynamic speckle classification in Advances in Self-Organizing Maps, Springer Berlin Heidelberg pp: 163–171.

[4] A. Zdunek, A. A. Damiak, M. Pieczywek, A. Kurenda (2014) Thebiospeckle method for the investigation of agricultural crops:A review. *Optics and Lasers in Engineering* (52) pp-276–285.

[5] S. Matharage, H. Ganegedara and D. Alahakoon, "A scalable and dynamic self-organizing map for clustering large volumes of text data," The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, 2013, pp. 1-8.

[6] Carlos Eduardo Bedregal Lizárraga C. E. (2008), Agrupamiento de Datos utilizando técnicas MAM-SOM, Tesis professional, UCSP - Universidad Católica San Pablo.

[7] A. L. Dai Pra; , G.Meschino, M. N.Guzmán; A. G. Scandurra, M. A. González; C. Weber; M.Trivi, H. Rabal, and L. I.Passoni (2016) Dynamic Speckle Image Segmentation *Journal of Optics* (18) No. 8.

[8] L. I. Passoni; H. J. Rabal; G. Meschino; and M. Trivi (2013) Probability mapping images in dynamic speckle classification *Applied Optics* No. 52 pp. 726–733.

[9] Braga Jr. R, Silva B, Rabelo G, Marques R, Enes A,; Cap N, Rabal H,Arizaga R, Trivi M, Horgan G. Reliability of biospeckle image analysis. Optics and lasers in engineering. 45(3) 390-95, 2007.

[10] Okamoto T, Asakura T. The statistics of dynamic speckles. Prog. Optics, 1995; 34. 183-248.

[11] J. D., Briers, Laser speckle contrast imaging for measuring blood flow OpticaApplicata, (2) No. 1–2, 2007.

[12] Rabal, H. J. and Braga, R. A. Dynamic Laser Speckle and applications (eds.) Boca Raton, FL, USA: CRC Press, 2008.

[13] J. B. MacQueen, (1967) Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* pp. 281–297.

[14] R.A. Braga, W.S. Silva, T. Safadi Time history speckle pattern under statistical view, C.M.B. Nobre. Optics Communications, 281, 2443–2448, 2008.

[15] R.A. Braga; W.S. Silva; T. Safadi (2008) Time history speckle pattern under statistical view, C.M.B. Nobre. *Optics Communications* No.281, pp.2443–2448.

[16] W.H. Payne; J.R. Rabung; T.P. Bogyo (1969). Coding the Lehmer pseudo-random number generator. *Communications of the ACM*. (12) No. 2 pp. 85–86.

Detection of Faults in WSNs based on NMF

Jimmy Ludeña-Choez, Juan J. Choquehuanca-Zevallos, Efrain Mayhua-López

Electronics and Telecommunications Engineering Research Center,
Universidad Católica San Pablo, Arequipa, Perú
jludenac, jchoquehuancaz, emayhua@ucsp.edu.pe

Abstract

Nowadays, Wireless Sensor Networks (WSN) are widely been employed for monitoring agriculture lands and so get useful information for efficient use of water resources. However, sensor nodes suffer from degradation producing erroneous measurements. In this paper, a machine learning method based on Non-Negative Matrix Factorization (NMF) is applied to the spectral representation of data stream from a WSN to model normal behaviour of the sensor nodes; and, by this means, detect faults in sensor nodes. Experiments on soil moisture data show that NMF achieves good results detecting flaws in readings from sensors.

1 Introduction

In these days, the development of methods to help to monitor proper usage of water resources for agriculture purposes is a fundamental task since the scarcity of water due to several reasons such as climate change or improper use made by man [6]. To this end, Wireless sensor networks (WSN) thanks to their nodes equipped with several types of sensors help to measure temperature, soil moisture, fertility, etc. [4, 5].

Unfortunately, WSN nodes are prone to failure due to a hostile environment, sensor aging, battery drain, human destruction, etc. In those cases, sensor nodes generate incorrect measurements that in turns lead to inappropriate decisions. This is the reason of developing efficient algorithms for fault detection and also to determine which sensors are the flawed ones.

Among the proposed methods for the Fault Detection task under these scenarios, PCA has extensively been used to detect faults by making a comparison between a learned model for normal behaviour and the data from sensor nodes. These PCA-based proposals have good performance due to the fact that measurements from sensors are highly correlated, providing in this way redundancy that is beneficial for the PCA algorithm [3]. Also, more elaborated methods such as the KPCA [1] seek to learn non-linearities existing in data. Other methods, for instance, the MSPCA [7] combine PCA with Discrete Wavelet Transform (DWT), allowing capture time-frequency information.

In recent years, the Non-Negative Matrix Factorization (NMF) algorithm has widely been applied for successful feature extraction, EEG (Electroencephalogram) signal processing, acoustic event classification (AEC), etc. [2, 8]. Basically, NMF unsupervisely decompose the data into a set of basis vectors and coefficients weighting such vectors. A better interpretability is achieved since the obtained representations are based on pieces of positive vectors. In this paper, NMF is used to model the normal behavior of the sensor nodes and thus to detect faults in soil moisture measurements transmitted from sensor nodes.

This paper is organized as follows: Section 2 presents the proposed front-end for the sensor nodes fault detection. Also, NMF is briefly presented. Section 3 presents the experimental setup and results. Finally, conclusions of this work are presented in Section 4.

2 Fault sensor nodes detection based on NMF

This section presents the procedure to detect faulty sensor nodes using NMF. To do that, we start with a brief description of the NMF method. Given a matrix $\mathbf{D} \in \mathbb{R}_+^{F \times N}$, where each column is the magnitude spectra of sensed data, NMF approximates it as a product of two non-negative low-rank matrices $\mathbf{W} \in \mathbb{R}_+^{F \times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K \times N}$ with $K \leq min\,(F, N))$ *i.e.* $\mathbf{D} \approx \mathbf{WH}$. In this way, each column of $\mathbf{D}$ can be written as a linear combination of K spectral basis vectors (SBVs) contained in the columns of $\mathbf{W}$, weighted with the coefficients of activation –or gain– located in the corresponding column of $\mathbf{H}$.

The factorization is achieved by an iterative minimization of a given cost function such as the KL divergence (Eq. 1). It results in the iterative update rules shown in Eq. 2 [2, 8].

$$D_{\mathrm{KL}}\left(\mathbf{D}\|\mathbf{WH}\right) = \sum_{ij}\left(\mathbf{D}_{ij} log \frac{\mathbf{D}_{ij}}{\left(\mathbf{WH}\right)_{ij}} - \left(\mathbf{D} - \mathbf{WH}\right)_{ij}\right) \tag{1}$$

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\frac{\mathbf{D}}{\mathbf{WH}}\mathbf{H}^T}{\mathbf{1}\mathbf{H}^T} \qquad \mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T\frac{\mathbf{D}}{\mathbf{WH}}}{\mathbf{W}^T\mathbf{1}} \tag{2}$$

where $\mathbf{1}$ is an $F \times N$ all-ones matrix. Multiplications $\otimes$ and divisions are component wise operations.

The faults in sensor nodes detection process is divided into two stages. Firstly, the normal behavior pattern of sensors is modeled by finding the SBVs by means of applying NMF to the magnitude spectrum of data. Then, all SBVs are concatenated to form a new matrix $\mathbf{W}_s \in \mathbb{R}_+^{F \times KS}$, where S is the total number of sensors.

Secondly, the detection of faulty sensor nodes is conducted through the calculation of the activation coefficients $\mathbf{H}_s$, such that $\mathbf{D}_{test} \approx \mathbf{W}_s\mathbf{H}_s$. To do so, the procedure depicted by Eqs. 1- 2 is performed again to approximate $\mathbf{D}_{test}$ but fixing $\mathbf{W}_s$ and only updating $\mathbf{H}_s$. Then for every sensor node, a characteristic vector is calculated as follows:

$$\mathbf{g}_s = \underset{n}{argmax}\left(H_{k_s}\right)_n, s \in \{1, ..., S\} \tag{3}$$

From these vectors, $\mathbf{G}_{test}$ matrix is constructed by concatenating the characteristic vectors of the each sensor ($\mathbf{g}_s$). Finally, the average gain matrix ($\mu_{test} = \sum_{k=1}^{K} G_{test}^{(k)}/K$) is used for monitoring sensors. If μ_{test} exceeds a threshold δ_{NMF} (calculated using the activations of the training data set as in Eq. 4), then the system determines that a sensor node has failed.

$$\delta_{NMF} = \sum_{s=1}^{S}\sum_{k=1}^{K} G_{train,s}^{(k)}/KS \tag{4}$$

where the $\mathbf{G}_{train}$ matrix is found by following the same procedure for obtaining the $\mathbf{G}_{test}$ matrix.

3 Experiments and results

3.1 Database and Baseline System

A database with agricultural soil moisture measurements was formed by collecting data from $S = 8$ sensor nodes. The measurements were taken every 5 minutes forming a total of 2000 samples for each sensor. Then, the dataset was divided into a training (first 1900 samples) and test (last 100 samples) sets.

The spectrum of data was computed using a Hamming window of 20 samples long with 50% of overlap. Regarding the faults of sensors, in this work, three types of faults have been experimentally induced in some sensors. They are Offset, Gain and Precision degradation fault.

3.2 Results

Results in Table 1 shows mean values for True Positive Rates (TPR), False Positive Rate (FPR) and Mean Error Probability (p_S) for 1000 experiments using three fault types with $Q_S = 1, 3, 7$ faults in readings from sensor nodes. It is worth to mention that the performance of NMF-based method seems to be more homogeneous for different faults, unlike the MSPCA-based method where a degradation of performance occurs when the number of faults increases. This can be observed as an increment in the value of p_S due to the rising in the probability of false alarm or FPR. This improvement was achieved primarily because the NMF-based method discovers the most important spectral bands that adequately represent the normal behavior of the sensor nodes.

Table 1: *Performance of NMF and MSPCA detection method for different faults in sensor nodes.*

Fault type		Precision degradation			Gain			Offset		
Faults in sensor nodes (Q_S)	Method	TPR (%)	FPR (%)	p_S (%)	TPR (%)	FPR (%)	p_S (%)	TPR (%)	FPR (%)	p_S (%)
1	NMF	90.30	0	4.85	82.80	0	8.60	87.00	0	6.50
	MSPCA	100	0	0	96.40	0	1.80	82.70	0	8.65
3	NMF	93.03	0	3.49	81.60	0	9.20	87.87	0	6.15
	MSPCA	99.80	0	0.10	71.07	1.38	15.16	76.57	0	11.72
7	NMF	90.97	0	4.51	81.60	0	9.20	88.50	0	5.75
	MSPCA	100	62.00	31.00	79.47	52.30	36.42	93.36	31.50	19.07

4 Conclusions

In this paper, a system for detecting faults in sensor nodes based on NMF was presented. The system models the normal behavior of the sensor nodes from the spectral basis vectors (SBVs) obtained after applying NMF over the spectral magnitude of the sensed data. From the SBVs, the activation coefficients in $\mathbf{H}_s$ are updated. Then, the sensor nodes are monitored using this gain matrix, enabling the system to determine which sensor nodes present wrong readings. The front-end has been tested using three types of faults artificially added to the test dataset. The results show that NMF algorithm is a promising tool to be used to detect flaws in readings from sensor nodes, allowing to capture the most important and relevant spectral components from sensed data.

References

[1] Oussama Ghorbel, Mohamed Abid, and Hichem Snoussi. Improved KPCA for outlier detection in Wireless Sensor Networks. In *1st Intl. Conf. on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 507–511, 2014.

[2] Jimmy Ludena-Choez and Ascension Gallardo-Antolin. Acoustic event classification using spectral band selection and Non-Negative Matrix Factorization-based features. *Expert Systs. with Applications*, 46:77–86, 2016.

[3] Harkat Mohamed-Faouzi, Djelel Salah, Doghmane Noureddine, and Benouaret Mohamed. Sensor fault detection, isolation and reconstruction using nonlinear principal component analysis. *Intl. J. of Automation and Computing*, 4(2):149–155, 2007.

[4] Tsuneo Nakanishi. A generative wireless sensor network framework for agricultural use. In *Makassar Intl. Conf. on Electrical Engineering and Informatics (MICEEI)*, pages 205–211, 2014.

[5] S Nandurkar, V Thool, and R Thool. Design and development of precision agriculture system using wireless sensor network. In *First Intl. Conf. on Automation, Control, Energy and Systems (ACES)*, pages 1–6, Feb 2014.

[6] UN-WATER. Coping with water scarcity - challenge of the 21st century. Technical report, United Nations, March 2007.

[7] Xie Ying-xin, Chen Xiang-guang, and Zhao Jun. Data fault detection for wireless sensor networks using multi-scale PCA method. In *2nd Intl. Conf. on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pages 7035–7038, 2011.

[8] Wang Yu-Xiong and Zhang Yu-Jin. Nonnegative Matrix Factorization: A comprehensive review. *IEEE Trans. on Knowl. and Data Eng.*, 25:1336–1353, 2013.

Analysis and selection of areas through clustering techniques for the Agropolis formation in Santander Colombia

Leonardo Hernán Talero-Sarmiento[1], Edwin Alberto Garavito-Hernandez[2], Henry Lamos-Diaz[3], Daniel Orlando Martinez-Quezada[4]

[1] Industrial University of Santander
Cll 9 # 27 Bucaramanga, Santander. Edificio Escuela de Estudios Industriales y Empresariales, of. 202
leonardo.talero@correo.uis.edu.co
[2] garavito@uis.edu.co
[3] hlamos@uis.edu.co
[4] daniel.martinez5@correo.uis.edu.co

Abstract

The agricultural production process begins with the type of land to be grown and its influence on productivity yield. Therefore, it is necessary to make a correct selection in order to avoid over costs or future losses. In the case of the Government initiative "Santander-Agropolis", it is essential to carry out an efficient selection of the potential sites. Thus, the objective of this work is to develop a tool for supporting localization decision making by applying clustering techniques focusing on the productive yield. The main results indicate that there are six groups that can be distributed in the six "Agropolis".

1 Introduction

The world hunger elimination is one of the United Nations sustainability goals, which include ensuring food security by promoting sustainable agriculture, strengthening rural development and protecting the environment. For this, Colombia as a United Nations member, decree in its development plan the strengthening of two axes related to this goal: *Development of a competitive rurality with an emphasis in the agricultural sector* and Environmental sustainability elements for rural development [2] seeking to generate well-being to the neediest population.

The reinforcement of the aforementioned axes becomes more relevant with the *FARC Peace Accord* signed in 2016, where different members of Governmental, Social, Academic and Industrial Institutions are invited to make efforts in order to generate a positive impact on the agriculture sector development [6], especially, focused on the well-being of small-scale farmers who own Agricultural Production Units (UPA) of less than five hectares [3] and had difficulty generating wealth through traditional processes.

Santander department is one of the main agricultural producers, however, in it has been identified an internal demand lack [4, 5], and small-scale farmer poverty; this is due to the concentration of power (and land) caused by the armed conflict and the incompatibility of production between agribusiness and food producers. In consequence, the Government seeks to generate strategies that balance both productive fronts (agro-industrial and food) along with increasing food safety, strengthening the small farmers and protecting the environment.

The Santander Governor proposed the formation of seven Agropolis which are models of agricultural associativity, focused on diversified sowing to meet the nutritional requirements of a population at a specific site, improving the quality of life of small-scale farmers, thus improving food resilience [5]. The Agropolis formation in each macro-region [4] is a development strategy which began with the formulation of the "*Santander-Magdalena Medio Agropolis*" in the macro-region called Mares and requiring to locate other six Agropolis in the remaining six macro-regions.

However, considering that the Agropolis seek to generate productive strategies integrating the various stakeholders and managing the agricultural production programming in a specific region [1], identify the appropriate area to carry out the project is necessary in order to facilitate the stakeholder's identification and make better use of arable land especially.

Moreover, taking into account the decision methods used by government are data lack (the methods are usually qualitative techniques based on the agreement between stakeholders), is necessary create tools to support decision activities with replicable techniques such as the presented in this work.

Therefore, the objective of the present research is to develop an analysis of the similarity and potential of several locations in Santander, Colombia, using a census data, applying machine-learning techniques, and generating a map. It indicates according to affinity the best sites to develop the six Agropolis which will strengthen the balance strategies proposed by the Governor of Santander within the framework of sustainable development objectives.

2 Methodology

The generation of groups was taken from features describing the production volume per unit area a summary of these variables is presented in Table 1. In order to determine the number of optimal groups for the dataset is used in the elbow method (see Figure 1) resulting in a $k = 6$. Then, the distance metric and k-means variant appropriate for the data set is used the sum of squares between relative groups (see Table 2), the selected combination corresponds to the one that has greater separation between groups corresponding to the combination of distance-algorithm Floyd and Euclidean.

	Tonnes/Hectare							
Agroindustry	Coffee	Palm Oil	Sugarcane	Cocoa	Tobacco	Cotton	Other	General
Tubers	Banana	Yucca	Potatoes	Other				General
Fruits	Banana special	Banana common	Critics	Papaya	Pineapple	Avocado	Other	General
Cereals	Rice	Yellow corn	White corn	Other				General
Vegetables	Vegetables	Species	Forest					General

Table 1: Variables Summary

Algorithm\Distance	*Euclidean*	*Maximum*	*Manhattan*
Hartigan	0.75	0.68	0.75
Lloyd	**0.78**	0.65	0.68
Forgy	0.72	0.69	0.74
Macqueen	0.71	0.58	0.52

Table 2: squared sum between clusters

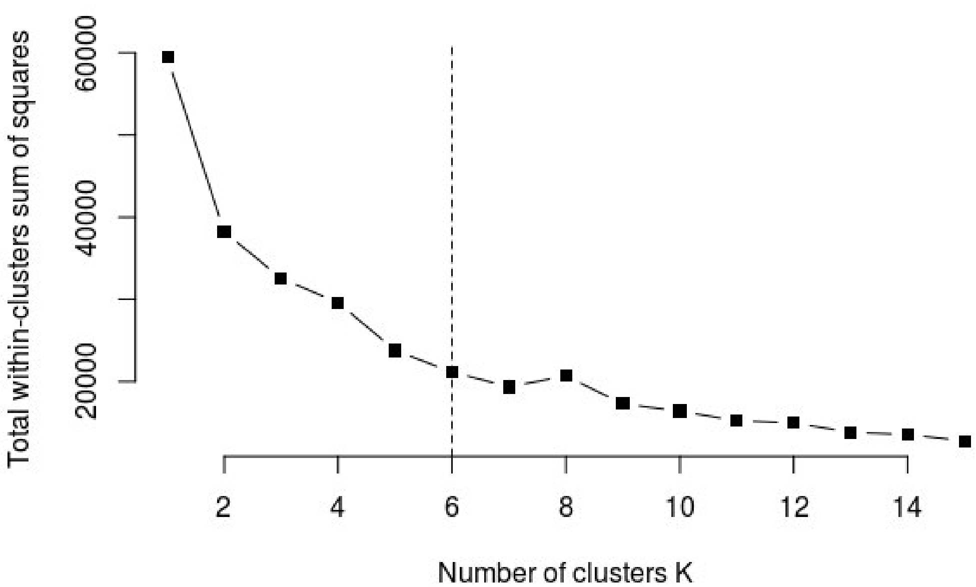


Figure 1: Elbow method for the optimum k

3 Results and conclusions

From the machine learning technique results, six types of clusters are formed. (See figure 2), Cluster 1 (light blue) has eighteen Townships (T) and is related to the higher food industrial productivity, Cluster 2 (green) is the most efficient tuber, potatoes, yucca, and banana producer with six T. Cluster 3 (orange) represents a group of varied crops with special efficient cereals sowing, and this group has seven T. Cluster 4 (purple) has presence in seven of eight macro-regions with fifteen T and is characterized by being the second best producer of fruits in Santander.

Cluster 5 (red) has one T called Lebrija and is the main pineapple producer in Colombia and the most efficient corn and cereal grower. Finally, Cluster 6 (dark blue) is the most massive group with forty Townships (for about 45.98% T), has a presence in the eight macro-regions and is the most efficient vegetables and fruits grower. The numbers of townships in each Cluster and macro-region is written in Table 3.

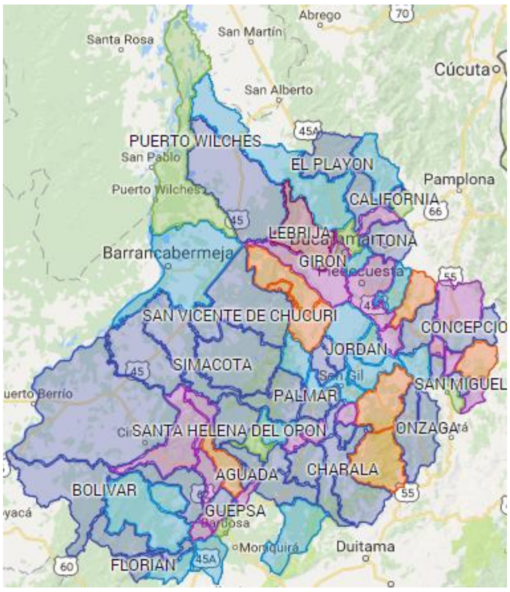


Figure 2: Clusters geospatial distribution in Santander

	Cluster					
Macro-region	Nº 1	Nº 2	Nº 3	Nº 4	Nº 5	Nº 6
Carare				1		2
Comuneros	3	1		2		9
García Rovira	1	1	2	4		4
Guanentá	4		2	2		10
Mares	2	1				3
Santurbán	1	1		1		3
Soto	3	1	2	2	1	1
Vélez	4	1	1	3		8

Table 1: Townships number in each Cluster and macro-region

From the results it is concluded that depending on the type of Agropolis (taking into account the product groups: Agroindustry, Tubers, Fruits, Cereals, and Vegetables) it is possible to select more than one region except in the macro-region called Soto, which has one T specialized in pineapple fruits citric fruits and corn-soybean, furthermore, taking into account the Townships geospatial distribution in each macro-region the Agropolis could be located in an area across many Municipalities increasing its food diversity.

References

[1] Ardila Hermano and Vergara Wilson, El modelo de agrópolis frente a la dialéctica ciudad-campo, *Rev. la Univ. la Salle*, vol. 57, no. 1, pages 83–95, 2012.

[2] Departamento Nacional De Planeación, Misión para la transformación del campo Colombiano, *15 de Enero de 2016 11:54 am*, 2015. [Online]. Available: https://www.dnp.gov.co/programas/agricultura/Paginas/mision-para-la-transformacion-del-campo-colombiano.aspx. [Accessed: 25-Apr-2016].

[3] Departamento Nacional De Planeación, El campo colombiano: Un camino hacia el bienestar y la paz. Tomo III. Bogotá: ISBN: 9789588340920, 2015.

[4] Gobernación de Santander, Plan de Desarrollo Departamental. Gobernación de Santander, Bucaramanga, page 230, 2016.

[5] Luc J.A., Agropolis: The Social, Political and Environmental Dimensions of Urban Agriculture. IDRC: Available: https://www.idrc.ca/en/book/agropolis-social-political-and-environmental-dimensions-urban-agriculture: ISBN: 1844072320, 2005.

[6] Oficina del alto comicionado para la paz, El acuerdo final de paz. La oportunidad para construir paz. Bogotá: Oficina del alto comicionado para la paz: Available: http://www.altocomisionadoparalapaz.gov.co/herramientas/Documents/Nuevo_enterese_version_6_Sep_final_web.pdf, 2016.

[7] Secretaría de Planeación Gobernación de Santander, Santander 2030 Síntesis del diagnóstico territorial de Santander. Bucaramanga, pages 1–250, 2011.

[8] Talero Sarmiento Leonardo Hernán, Rodriguez Torres Leidy Tatiana and Diaz Bohorquez Carlos Eduardo, Definición y caracterización de los principales productos agrícolas cultivados en las provincias de García Rovira y Guanentá; como base para la propuesta de Primera Milla, Bucaramanga, 2016.

Metaheuristic algorithms for multi-objective optimization in dairy systems

Gastón Notte[1], Martín Pedemonte[2], Héctor Cancela[2], Pablo Chilibroste[3]

[1] Centro Universitario de Paysandú, Universidad de la República
Paysandú, Uruguay
notteg@cup.edu.uy

[2] Facultad de Ingeniería, Universidad de la República
Montevideo, Uruguay
mpedemon@fing.edu.uy, cancela@fing.edu.uy

[3] Facultad de Agronomía, Universidad de la República
Montevideo, Uruguay
pchili@fagro.edu.uy

Abstract

The dairy industry is very important for the Uruguayan economy, and it presents many opportunities for attaining better efficiency levels by using operational research techniques. In this paper we address the problem of food resources allocation in pastoral based dairy systems, which consists of determining how to distribute the available resources to the herd.

The main goal was to develop a multiple objective optimization model, covering multiple periods and integrating operational decisions solved by more detailed submodels. To find solutions for this model, we programmed an evolutionary algorithm and studied the best parameter configuration to obtain a good computational performance.

1 Introduction

Dairy industry is one of the most complex and important sectors of the Uruguayan economy, and the production has been increased in the last decades. In particular, in the last 7 years the milk production has been growing at rates of 7% per year [6], and the milking area has been reduced by 20%. There was a significant increase in productivity per hectare and productivity per cow, and this increase is mainly due to competition with other agricultural activities and because of higher land prices [2].

Based on the data presented by DIEA [6], in 2007 production reached 750 liters per hectare, while in 2014 reached 4000 liters per hectare. However, these values represent very variable situations, while some producers increased their production by 4% other producers grew at a rate of 13%.

In Uruguayan dairy production there are 3610 producers, using a total area of 762000 hectares. In total they have 440000 cows and generated a production of 2130 million liters of milk in 2014 [6].

The dairy production systems in Uruguay are defined as pastoral systems with supplementation [3]. In these systems the stocking rate is the main factor that determines the effectiveness of the system, directly impacting on milk production and forage consumption [2]. The intensification of milk production in Uruguay is based on a significant increase in the use of concentrates and conserved forage [5], while direct forage harvesting by the animals has remained unchanged [3]. However, the viability of these practices and their productive and economic sustainability is not very clear. The intensification of milk production systems represents a strong impact on the entire dairy chain due to a lower area requirement and the possibility of higher prices of raw material.

Because of the importance of the dairy industry for the Uruguayan economy, the complexity of the dairy management systems and the increase of the intensification process, it is of high interest to study problems related to dairy systems using an operational research focus to enrich traditional agronomy approaches. Particularly, in this work we address the problem of food resources allocation in pastoral based dairy systems.

2 Problem description and solution method

The food resource allocation to a dairy herd consists in determining how to distribute the available resources considering different objectives. Those resources are different types of food located in field areas that must be allocated to the cows. In Uruguay, the resource allocation is done by dividing the herd into groups of cows (each group can have different sizes and include animals of different characteristics), and then distributing those groups into different feeding areas. In general, to simplify the food allocation task, each group remains unchanged (same cows) for a certain period of time. There are many combinations on how to group the animals; and even more combinations can be considered in order to assign those groups to the existing resources, with some solutions being much better than other ones. In Uruguay, this allocation process is usually based on the experience, intuition (and even traditions) of the producers, following some management rules considering parturition, days in milk, actual milk production, among others.

This type of allocation can be addressed by a combinatorial optimization model. Trying to optimize dairy systems is hard and complex, specially because there are many factors to consider (attributes, objectives and constraints), but a great advantage of this approach is to have the opportunity to help farmers and producers to explore a large number of combinations and finally follow the solution that fits better for them.

The optimization techniques are widely considered as very useful for agricultural models. Mathematical programming methods were among the first used approaches for agricultural optimization, and then, many studies using linear programming methods have been published [14]. Also, different papers using metaheuristics for land use optimization have been reported [11].

Dairy production is also an area where optimization techniques have been used [4, 7–10, 13].

In general, the problems referenced above did not consider the animal grouping and did not differentiate how cows of different types were fed. A first approximation of that problem was presented by Notte [12], and was presented in terms of supply and demand. The supply structure was defined by the availability of food resources and their characteristics, while the demand structure was defined by the energy required by the herd (based on the nutrient requirements of dairy cattle as published by the NRC). The model in that work is a single objective model, and defines new groups of cows for every single milking, thus resulting in solutions which are not practicable for Uruguayan producers.

Considering the difficulty of large-scale optimization problem, where traditional exact approaches cannot be applied, a very good alternative are metaheuristics, which have been used to obtain good quality approximate solutions in a reasonable execution time. Evolutionary Algorithms [1] have proven to be flexible and robust methods for effectively solve complex optimization problems.

In this work, the main goal was to develop a more realistic model for determining how to allocate the available resources. A mathematical programming model was constructed, taking into account the characteristics of the problem. The model follows a multi-period approach, covering a one year schedule, divided into periods of one month. Decision variables correspond to how to divide the entire herd into groups of animals, which remain in place during each one mont period (but can be changed from one month to the following one). Parameters of the model include the different food types and availabilities, as well as the characteristics of each cow in the herd. The intra-month behavior of the system is represented using more detailed models (in particular, taking into account the changes in grass availability due to the consumption by the herd and the vegeal growth; and also taking into account the milk production of each cow, depending on the food availability via the decided allocation, and depending on genetics and other characteristics).

Finally, to solve the integrated optimization model we developed a multi-objective evolutionary algorithm for the optimization problem, taking as decision variables a sub-set of input parameters, and using different objective functions (maximizing total production, maximizing production efficiency, maximizing economic gain, minimizing capital invested, etc.). We also studied the best parameter configuration in order to obtain a good computational performance.

The results obtained showed that it was possible to find in a reasonable computing time a number of good quality solutions, representing different tradeoffs between the objective functions mentioned above, and approximating the Pareto front of efficient decisions. This set of good quality solutions can be used by the farmers, to select an efficient food allocation strategy that corresponds to a particular tradeoff suiting their own personal preferences and constraints.

References

[1] T. Bäck. *Evolutionary algorithms in theory and practice: Evolution strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, 1996.

[2] P. Chilibroste. Carga o Productividad Individual?. Pasto o concentrado?: mitos y realidades en la intensificación de los sistemas de producción de leche en Uruguay. *Editorial: Centro Medico Veterinario de Paysandu*, 1:158–162, 2015.

[3] P. Chilibroste, P. Soca, and D. A. Mattiauda. Estrategias de alimentación en Sistemas de Producción de Leche de base pastoril. En: Pasturas 2012 : Hacia una ganadería competitiva y sustentable. *Balcarce: INTA.*, pages 91–100, 2012.

[4] G. Dean, H. Carter, H. Wagstaff, S. Olayide, M. Ronning, and D. Bath. Production functions and linear programming models for dairy cattle feeding. *Giannini Foundation of Agricultural Economics, University of California*, 31:1–54, December 1972. [Online]. http://giannini.ucop.edu/Monographs/31-DairyCattleFeedingModels.pdf.

[5] DIEA. La producción lechera en el Uruguay. Technical report, Estadístico Agropecuario 2009. Serie de encuesta Nro 278 Montevideo; MGAP. 79p., 2009.

[6] DIEA. Anuario Estadístico Agropecuario. Technical report, Montevideo: MGAP. 243p., 2014. [Online]. http://www.mgap.gub.uy/Dieaanterior/Anuario2015/DIEA-Anuario-2015-web.pdf.

[7] G. Doole and A. Romera. Detailed description of grazing systems using nonlinear optimisation methods: A model of a pasture-based New Zealand dairy farm. *Agricultural Systems*, 122:33–41, November 2013.

[8] G. Doole, A. Romera, and A. Adler. A mathematical model of a New Zealand dairy farm: The Integrated Dairy Enterprise Analysis (IDEA) framework. Working Paper 1201, Waikato University Department of Economics, Hamilton, New Zealand, 2012. [Online]. ftp://mngt.waikato.ac.nz/RePEc/wai/econwp/1201.pdf.

[9] G. Doole, A. Romera, and A. Adler. An optimization model of a New Zealand dairy farm. *Journal of Dairy Science*, 96(4):2147–2160, April 2013.

[10] A. Kalantari, H. Mehrabani-Yeganeh, M. Moradi, A. Sanders, and A. De Vries. Determining the optimum replacement policy for Holstein dairy herds in Iran. *Journal Dairy Science*, 93(5):2262–2270, May 2010.

[11] M-M. Memmah, F. Lescourret, X. Yao, and C. Lavigne. Metaheuristics for agricultural land use optimization. a review. *Agronomy for Sustainable Development*, 35(3):975–998, 2015.

[12] G. Notte, M. Pedemonte, H. Cancela, and P. Chilibroste. Resource allocation in pastoral dairy production systems: Evaluating exact and genetic algorithms approaches. *Agricultural Systems*, 148:114 – 123, 2016.

[13] B. Ridler, J. Rendel, and A. Baker. Driving innovation: Application of linear programming to improving farm systems. In *Proceedings of the New Zealand Grassland Association*, pages 295–298, 2001.

[14] A. Weintraub and C. Romero. Operations research models and the management of agricultural and forestry resources: A review and comparison. *Interfaces*, 36(5):446–457, 2006.

Simulated Annealing in the Operational Forest Planning

Paulo Amaro Velloso Henriques dos Santos[1], Arinei Carlos Lindbeck da Silva[2], Julio Eduardo Arce[3]

[1] Instituto Federal de Santa Catarina – Campus Joinville
Rua Pavão, 1377 – Costa e Silva, Joinville/SC – Brasil
pauloamaro.ifsc@gmail.com

[2] Universidade Federal do Paraná – Campus Politécnico
ACF Centro Politécnico, Jardim das Américas, CEP: 81531-980, Curitiba/PR – Brasil
arineicls@gmail.com

[3] Universidade Federal do Paraná – Campus Jardim Botânico
Av. Prof. Lothário Meissner, 900 - UFPR – Bloco B3, CEP 80210-170, Curitiba/PR – Brasil
jarce@ufpr.br

Abstract

This paper proposes a method to optimize the Operational Forest Planning based on the Simulated Annealing metaheuristic. The Operational Forest Planning is the hierarchical level of the forest planning that encompasses the harvesting and the transportation of timber. The proposed method aims to find a lower cost solution for harvesting and transportation of timber to the demand centers. The results show that the method can improve the first feasible solution found from 30% to 68%.

1 Introduction

By looking at the amount of papers published in the last decades, we can understand that the optimization of forest planning is necessary, since this research field moves high values and it serves several clients, such as paper industries, furniture industries, energy production sectors, among others. Since the early '90s the amount of papers published about forest planning increased especially in strategic and tactical levels. The operational level presents the smallest amount of papers at the moment.

The authors were not able to find another paper with the same focus on optimization that they had in this paper, so it was difficult to compare the results of this paper with another reference. The authors made use of constructed scenarios based on real scenarios and compared the best solution found with the first feasible solution. The authors aim to show that the proposed method can be a great reduction in the cost of operational forest planning by spending relatively small computational time.

2 Forest Planning

2.1 Operational Forest Planning

The papers and literature published by the end of the '80s show the optimization of forest planning through performing this optimization in seeds, aiming to maximize the harvest. Between the late '80s and the early '90s, researches began to use different hierarchical levels in forest planning: Strategic, Tactical and Operational. [2] published one of the first papers showing these different levels in their development.

The Operational Forest Planning is the hierarchical level of the forest planning that encompasses the harvest and the transport of wood. The operational level currently presents less papers than others levels, probably because this level needs much more variables than the others. Besides, it is necessary for the process not to take long because, this level has a short deadline.

The development of a method to optimize forest planning in the operational level presents high computational complexity and, in addition, we do not have real data of forest scenarios available to test and improve the method. Nowadays, the most used way for this development is to implement arrangements with private companies, using its forest data and trying to optimize the harvest for this company. [10] shows a tool to generate forest scenarios based on real data of a forest. That tool was used on the development of this paper to generate scenarios to test the proposed method of optimization.

2.2 Forest Scenarios

The forest scenarios created by the tool show all the necessary information to understand and optimize the operational forest planning in those scenarios. Each scenario presents the following information:

- Amount of plots ready to be harvested;
- Amount of different forest products that can be harvested on a plot;
- For each plot: the identification, the coordinates, the size, the topography, the specie of tree seeded, the conditions of harvest in rainy season, the inventory of each forest product available on this plot;
- Amount of demand centers, where the forest products will be delivered;
- For each demand center: the identification and the coordinates.

For the development of the method presented on this paper, the authors used ten different scenarios. These scenarios are available at www.joinville.ifsc.edu.br/~paulo.amaro/Forest_Scenarios. They are named "Cenário 1" to "Cenário 10".

3 Optimization Method

The optimization method chosen was based on the Simulated Annealing metaheuristics. It was chosen because for each solution made to the harvest problem it was necessary to solve a transportation problem for the forest products. The transportation problem is a complex problem to solve; Simulated Annealing demands less computational effort than other metaheuristics as Genetic Algorithm applied in forest problems. Many researches were published in last three decades using Simulated Annealing in forest problems such as [1], [3], [4], [6], [7] and [8], especially for solving spatial constrained harvest problem. The Simulated Annealing is a metaheuristic developed by [11], based in [9].

Besides the information about the scenario, it was necessary to input the planning horizon, the demands of each forest product and the information about the harvest teams. After these steps, the method creates random sequences of plots for each harvest team. Based on the sequence and the offers of products, the forest planning problem was transformed in a sparse transportation problem. The transportation problem was solved with the method presented by [5]. The method can find the optimal solution for the transportation problem. The cost of this problem was used to compose the total cost of the operational forest planning in addition to the costs for transportation and operation of the harvest teams. The total cost is the evaluation function for the Simulated Annealing and it can be represented by this equation:

$$F = \sum_{i,j,k,l,m} x_{ijklm} + \sum_{n,o,p,q} y_{nopq} + \sum_{r,s} z_{rs} \tag{1}$$

where x_{ijklm} is the transportation cost for the forest product i harvested in plot j in day k and transported to demand center l in day m, y_{nopq} is the cost for moving the harvesting team n in day o from the plot p to plot q and z_{rs} is the operational cost of harvest team r in day s.

The Simulated Annealing was used for 10 minutes. After this, the best solution found by the metaheuristic was assumed like the solution for the problem. Besides that, the method compared the best solution with the first feasible solution found and the difference in percentage between these two solutions was then calculated, called percent reduction of solution.

4 Tests and Results

Tests were done in order to evaluate the proposed method. Some parameters needed to be determined before the tests. The planning horizon was set in 7 days and the parameter of Simulated Annealing used to update the temperature was called alpha (represented by α) and three alpha values were determined for the tests (0.80, 0.90 and 0.95).

Ten tests were performed with each scenario for each alpha value. The results were compiled and some information by these tests are shown in Figures 1 and 2. Figure 1 shows the behavior of the average of percent reduction of the best solution found in relation to the first feasible solution found for each alpha value. It can be observed in Figure 1 that the best solution found was, at minimum, 30% better than the first feasible solution found previously. In addition, it can be observed that alpha 0.80 reaches better solution than other alpha values in 6 for 10 tested scenarios while alpha 0.90 reaches better solutions than other alpha values only in one tested scenario.

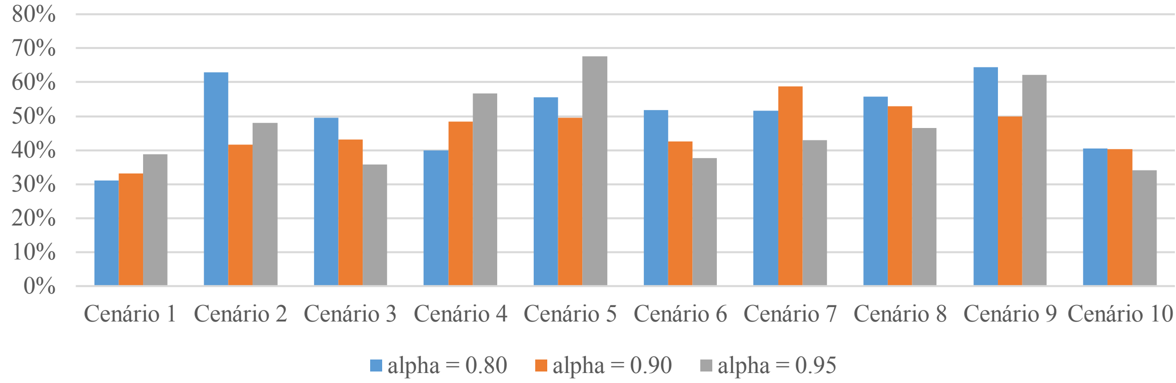


Figure 1: Graphic with the average of percent reduction of solution

References

[1] André O. Falcão and José G. Borges. Combining random and systematic search heuristic procedures for solving spatially constrained forest management scheduling models. *Forest Science*, 48:608–621, 2002.

[2] Andrés Weintraub and Alejandro Cholaky. A hierarchical approach to forest planning, *Forest Science*, 37:439–460, 1991.

[3] Carey Lockwood and Tom Moore, Harvest scheduling with spatial constraints: a simulated annealing approach. *Canadian Journal of Forest Research*, 23:468–478, 1993.

[4] Emin Z. Baskent and Glen A. Jordan. Forest landscape management modeling using simulated annealing. *Forest Ecology and Management*, 165:29-45, 2002.

[5] Gustavo V. Loch. Uma nova abordagem no processo iterativo de melhoria de solução no problema de transporte. Doctoral thesis, Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Brasil, 2014.

[6] Kevin A. Crowe and John D. Nelson. An evaluation of the simulated annealing algorithm for solving the area-restricted harvest-scheduling model against optimal benchmarks. *Canadian Journal of Forest Research*, 35(10):2500-2509, 2005.

[7] Kevin Boston and Pete Bettinger An analysis of Monte Carlo integer programming simulated annealing and tabu search heuristics for solving spatial harvest scheduling problems. *Forest Science*, 45:292–301, 1999.

[8] Lucas R. Gomide, Julio E. Arce and Arinei C. L. Silva. Comparação Entre a Meta-Heurística Simulated Annealing e a Programação Linear Inteira no Agendamento da Colheita Florestal com Restrições de Adjacência. *Ciência Florestal*, 23(2):449-460, 2013.

[9] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller and Edward Teller. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21: 1087–1092, 1953.

[10] Paulo A. V. H. dos Santos, Arinei C. L. da Silva and Julio E. Arce. Uma ferramenta para construção de Cenários Florestais, *Proceedings of the XVIII Latin-Iberoamerican Conference on Operations Research (CLAIO 2016)*, pages 819-826, Santiago, Chile, October 2–6 2016.

[11] Scott Kirkpatrick, Daniel Gelatt and Mario P. Vecchi. Optimization by Simulated Annealing. *Science*, 220:671–680, 1983.

Identifying trade-offs between sustainability dimensions in the supply chain of biodiesel in Colombia

Javier Arturo Orjuela-Castro[1], Johan A. Aranda-Pinilla[1], Carlos E. Moreno-Mantilla[2]

[1] Faculty of Engineering, Industrial Engineering Department, Universidad Distrital, Francisco José de Caldas
jaorjuelac@unal.edu.co, jaarandap@udistrital.edu.co

[2] Faculty of Engineering, Systems and Industrial Engineering Department, Universidad Nacional de Colombia
cemorenoma@unal.edu.co

Abstract

This paper proposes and develops a deterministic multi-objective linear programming model to analyze the relationship among the economic, environmental, and food security dimensions of the biodiesel supply chain (BSC) in Colombia. Considering four echelons from the supply chain (palm cultivation, oil extraction, bio-refineries, and mixers), the model seeks to minimize total cost of the chain, impact on food security, and emissions of greenhouse gases, including emissions from direct and indirect land use change. The Epsilon-constraint method is used to solve the multi-objective model for the BSC. A Pareto set of optimal solutions helped to identify trade-offs involving the three objectives.

1 Introduction

Global production of biofuels has grown steadily in recent years. However, despite their advantages over fossil fuels, there are concerns about the environmental and social impacts of biofuels' production and distribution. According to the Food and Agriculture Organization of the United Nations (FAO, 2008), one of the most important social issues in biofuels chains is the impact they can have on food security. Regarding environmental issues, a study prepared for the Inter-American Development Bank (IDB) and the Mines and Energy Ministry of Colombia highlights the impacts associated with direct and indirect land use change (CUE Consortium, 2012).

Consequently, this paper proposes and develops a multi-objective linear programming model to help understand the relationship between economic, environmental and food security performance indicators in the biodiesel chain for the case of Colombia. Considering four links in the supply chain (African palm cultivation, oil extraction, bio-refineries, and mixers), the model seeks to minimize three objectives: total cost of the chain, impact on food security, and emissions of greenhouse gases (GHG), including emissions from direct and indirect land use change.

2 Research methodology

After an analysis of the state-of-the-art models and techniques, the objectives for the management of the supply chain, and a diagnosis of the biodiesel supply chain (BSC) in Colombia, the relationship between cost, food security and GHG emissions was established. From these relationships the parameters and variables were identified and a multi-objective linear programming model was proposed. Later, experimentation with the model was applied to the case of the BSC, which considered crop yields throughout the palm's life cycle, projected production capacity, and projected demand in Colombia for a horizon of 30 years. Based on a scenario where the deficit of biodiesel at a nearly optimal cost per ton produced is reduced, the Epsilon-constraint method was used to generate Pareto frontiers that helped to identify transactional solutions that consider the economy, the environment, and food security. The model was programmed in GAMS.

3 Findings

The analysis of Pareto frontiers (see Figure 1) shows that there is an inverse relationship between the three objectives evaluated. Applying the lexicographical method, different objectives are achieved: in A, a minimum cost is achieved with a low level of emissions and a high impact on food security; in B, optimal emissions are maintained while the impact on food security (0 Ha-year replaced) is minimized; and, point C is obtained by keeping the level of costs from point A and minimizing the impact on food security, which results in an increase in emissions.

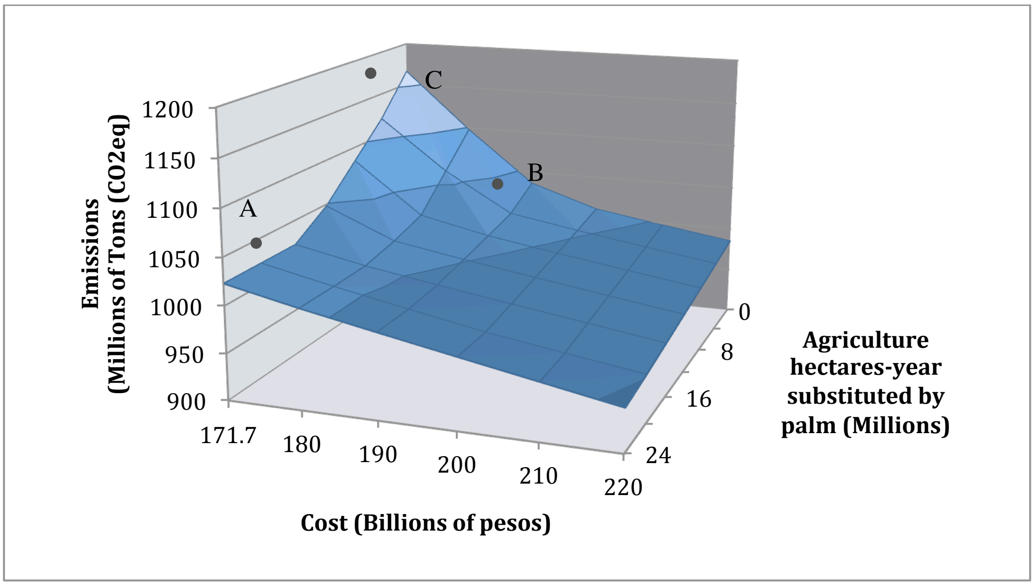


Figure 1 –Pareto frontiers for the relationships between sustainability objectives

The cost is inversely proportional to emissions since the most productive soils are the most emissions-intensive, as in the case of non-protected forests. In turn, using less productive soils as grasslands or scrubs increases absorption of carbon dioxide (CO_2) but at the expense of increased production costs. The inverse relationship between cost and food security can be explained in that there are soils dedicated to agricultural crops, which might have higher yields of palm per hectare than they would in grassland soils or scrub. The impact on food security can be reduced if the substituted crops are finally replaced in a natural area (forest or grassland), which would cause an increase in CO_2 emissions by changing land use.

4 Relevance and contribution

In modeling sustainable supply chains, trade-offs between the results in the environmental, social and economic dimensions are the rule, rather than the exception (Brandenburg et al, 2014; Seuring, 2013). However, empirical studies are required that support or reflect on how particular situations can generate these trade-offs (O'Rourke, 2014; Dekker et al, 2012; Wu and Pagell, 2011).

The model developed in this article has served to establish the relationship between the three objectives evaluated, which helps support decision-making in the BSC and guides the definition of sustainability-oriented policies. The proposed generic mathematical model can be applicable to any supply chain of agro-fuels and allows for the calculation of a production, inventory and distribution plan of raw materials, intermediate, and finished products throughout the entire chain.

In addition, the model has features not found in the review of the state of the art. First, the models incorporates variations in palm yields per hectare over the life of the crop, allowing to define an optimum planting production plan to achieve the levels required in the next 30 years. Also, the model considers GHG emissions per hectare in the growth phase (not per ton of fruit) and takes into account different types of soil, making explicit the variations in production and emissions depending on the type of soil used, including emissions from indirect land use change, thus explaining the trade-off between GHG emissions and food security.

References

[1] Brandenburg, M., K. Govindan, J, Sarkis, and S. Seuring (2014), "Quantitative models for sustainable supply chain management: Developments and directions", *European Journal of Operational Research*, Vol. 33, No. 2, pp. 299–312.
[2] Consorcio CUE (2012), *Evaluación del ciclo de vida de la cadena de producción de biocombustibles en Colombia*, Ministerio de Minas y Energía, Medellín.
[3] Dekker, R., J. Bloemhof, and J. Mallidis (2012), "Operations Research for green logistics – An overview of aspects, issues, contributions and challenges", *European Journal of Operational Research*, Vol. 219, No. 3, pp. 671–679.
[4] FAO (2008), *The state of food security in the world 2008*, FAO, Roma.
[5] O'Rourke, D. (2014), "The science of sustainable supply chains", *Science*, Vol. 344, No. 6188, pp. 1124–1127.
[6 Seuring, S. (2013), "A review of modeling approaches for sustainable supply chain management", *Decision Support Systems*, Vol. 54, No. 4, pp. 1513–1520.
[7] Wu, Z., and M. Pagell (2011), "Balancing priorities: Decision-making in sustainable supply chain management", *Journal of Operations Management*, Vol. 29, No. 6, pp. 577–590.

Body condition estimation on cows from 3D images using Convolutional Neural Networks

Juan Rodríguez Alvarez[3], Mauricio Arroqui[1,3], Pablo Mangudo[1,3], Juan Toloza[1,3], Daniel Jatip[1,3], Juan M. Rodríguez[2], Alejandro Zunino[2], Claudio Machado[1,3], Cristian Mateos[2]

[1] Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT)

[2] ISISTAN-UNCPBA-CONICET. Tandil, Buenos Aires, Argentina

[3] CIVETAN-FCV UNCPBA-CONICET-CIC. Tandil, Buenos Aires, Argentina

Abstract

BCS ("Body Condition Score") is a method to estimate body fat reserves and accumulated energy balance of cows. BCS heavily influences milk production, reproduction, and health of cows. Therefore, it is important to monitor BCS to achieve better animal response. It is a time-consuming and subjective task, performed visually by expert scorers. Several studies have tried to automate BCS of dairy cows by applying image analysis and machine learning techniques. This work analyzes these studies and proposes to use a method based on Convolutional Neural Networks (CNNs) to improve overall automatic BCS estimation and extending its use beyond dairy production.

1 Introduction

Body condition score (BCS) refers to the relative amount of subcutaneous body fat or energy reserve in cows, regardless of body weight and frame size [24]. BCS uses a 5-point scale with 0.25-point increments (ranging from 1 -emaciated cows-, to 5 -obese cows-) [8, 9]. BCS is an important management tool, which can improve herd nutrition, health, production and pregnancy rate [13, 15, 18, 19]. BCS estimation is a time-consuming process measured manually that requires trained evaluators. The subjectivity in the judgment of evaluators can lead to different scores for the same cow under consideration, or even inconsistent scores by the same expert. Thus, with the increasing advances in technology availability at an accessible cost, automation and digitalization of livestock farming tasks offer multiple opportunities. Different studies have particularly focused on BCS automation. This extended abstract identifies the most recent studies on this area and highlights some important aspects, depicting concrete opportunities for further development.

2 Related Works

Several attempts to automate the determination of dairy cows' BCS using digital images are reported in the literature. Developed methods have two stages: (i) image analysis techniques to extract relevant characteristics (such as angles, distances and areas between anatomical points; intensity/depth pixels values; cow contour or a representation of it) to differentiate fat reserves levels of cows; (ii) usage of collected characteristics to implement a BCS estimation model. Mostly, there are two types of models used: regression analysis models (as in [2, 4, 5, 10, 17, 20, 22]) and algorithms that measure cow's body angularity (as in [11, 12, 21]) according to the hypothesis that the body shape of a fatter cow is rounder than that of a thin cow.

Moreover, three automation levels exist. In the lowest level are [2, 5, 10], which require to manually identify anatomical points in the images to extract characteristics to develop the estimation models. In the medium level are [1, 4, 17], where the input images used are manually selected, but the rest of the process is automatic. Finally, in the highest level are [11, 12, 20, 21, 22], which achieve a completely automated process. Among the latter studies, only [11, 12] carry out real time estimations because image preprocessing techniques (segmentation, normalization, features extraction) used in the other studies are time-consuming and therefore are performed under a batch scheme. However, [11] use a very expensive thermal camera (in comparison with the other studies) and [12] do not perform a detailed analysis of results and only corroborate the inversely proportional relationship between BCS and angularity of the cow's body.

3 Opportunities

Despite some studies proposed automatic systems and achieved good BCS estimation results within the expected error range in comparison to expert's scores ([11, 21, 22]), none has simultaneously developed a highly automated, accurate, real-time and low cost method. Particularly, real time evaluation does not represent a problem on dairy farm activities because farmers interact with the cows at least twice a day. However, it is very important in beef cow-calf operations, where interactions with cows have seasonal frequency, and contingency actions should be applied immediately (e.g. herd split) to avoid unnecessary herd movements. In the same way, a system oriented to dairy and beef breeding operations needs a broad training and validation images set, involving different cow breeds. In contrast to surveyed studies, the new method needs taking into account different cow frames to achieve accurate estimations, using a cheap camera. These challenges open up opportunities for developing BCS estimation systems for dairy production and beyond.

Additionally, an alternative powerful machine learning technique from the field of deep learning, known as Convolutional Neural Network (CNN), has not been proven yet. CNNs [3] have been found highly effective and been commonly used in computer vision and image classification [7, 14, 16, 23]. A CNN is a specialized kind of neural network with a special architecture composed of a sequence of layers. Three main types of layers are used to build a CNN: convolutional, pooling (or subsampling) and fully-connected layers. In the traditional model of pattern/image recognition (studies of Section 2) a by hand-designed feature extractor gathers relevant information from the input image. Then, features are used to train a classifier (or a regression model), which outputs the class (or value) corresponding to an input image. In a CNN, convolution and pooling layers play the role of feature extractor, where the weights (model coefficients or parameters) of the convolutional layer being used for feature extraction as well as the fully connected layer being used for classification are automatically determined during the training process [14]. In this way, a CNN transforms the original image layer by layer from the original pixel values to the final class scores (the discrete BCS values within the 5-point scale).

Furthermore, when the BCS system starts to work in the farm, a huge number of images and body condition values will be periodically generated. This amount of data could be analyzed individually or together with other information sources, such as sensors around the farm and on animals (e.g. activity meter collar), local and external information systems, custom digital registers, etc. Thus, Big Data techniques could be applied to organize, analyze, process and interpret such large volumes of diverse data.

4 Current Status

A system to estimate BCS in real time is being developed. Images are being collected to build a dataset of cows with their associated BCS. Then, these images will be used to train and validate a CNN model.

4.1 Data Collection and Model Validation

Three dairy farms has been visited to acquire images. One of them is located in Carlos Pellegrini, Santa Fe (Argentina), and has about 1000 cows. Two are located in Gardey, Buenos Aires (Argentina) and have 200 and 400 cows, respectively. Figure 1 shows the device used to capture images while the cows walked voluntarily below the camera (Microsoft Kinect V2 ToF). This type of camera is raising interest in livestock application for its high quality and low cost (around U$S100). The device was located at the exit of milk parlor, 2.8m above ground and aimed downward. We will initially use depth 512x424 images to train/validate the model. Depth images have proved to be more suitable than RGB images to depict cow's body variability associated with changes in BCS [10].

During the acquisition of the cow images, an expert scorer evaluates in situ the BCS of cows to build a consistent labeled dataset. Cows were scored for the same expert in the three dairy farms to reduce subjectivity inconsistencies.

To date, the dataset built has around 1500 depth cows images. The number of necessary images to get good estimations results will depend on the CNN design and configuration (hyperparameters), and the difficulty of the learning problem, i.e. correctly distinguishing BCS values. However, a dataset of around 500 images per class (each possible BCS value), in combination with the use of data augmentation techniques, should be suitable.

The percentage of correct BCS model estimations within the range of human error (0.25 - 0.5, equals to one-two intervals/classes of distance) will be the principal validation measure, which is one the most frequently used approach in the literature. This will allow us to compare the obtained results against the other studies.

Figure 1: Device (see top) to capture images at a dairy farm.

4.2 Development Tools

The BCS estimation software is being written in Python, and Keras (`https://keras.io/`) is being used to develop the image classifier model using CNN. Keras is a model-level library that provides high-level building blocks for developing deep learning models, and works on top of Theano or Tensorflow. These frameworks serve as "backend engines" of Keras.

Keras models can run on GPU, thus speeding up training and inference by a considerable factor (often 5x to 10x, when going from a modern CPU to a single modern GPU). A GPU can perform lots of simple numerical processing task at the same time (massively parallelized), such as the huge amount of matrix multiplications and other relevant operations associated to CNN. Keras uses cuDNN [6] for high-performance GPU acceleration. cuDNN (part of the NVIDIA Deep Learning SDK) is a GPU-accelerated library that provides highly tuned implementations for standard CNN routines such as forward and backward convolution, pooling, normalization, and layer activation.

5 Conclusion

Despite automatic methods to estimate BCS are available, new (undergoing) development opportunities have been identified to implement an automatic, accurate, real-time, and low cost BCS estimation system, allowing its application beyond dairy production. The cornerstone of this system are CNNs, an effective technique to classify images, which could improve BCS estimations accuracy in relation to previous works.

References

[1] D. Anglart. *Automatic estimation of body weight and body condition score in dairy cows using 3D imaging technique*. Master's thesis, 2010.

[2] G. Azzaro, M. Caccamo, et al. Objective estimation of body condition score by modeling cow body shape from digital images. *Journal of Dairy Science*, 94(4):2126 – 2137, 2011. ISSN 0022-0302.

[3] Y. Bengio, I. J. Goodfellow, et al. Deep learning. *Nature*, 521:436–444, 2015.

[4] A. Bercovich, Y. Edan, et al. Development of an automatic cow body condition scoring using body shape signature and fourier descriptors. *Journal of dairy science*, 96(12):8047–8059, 2013.

[5] J. Bewley, A. Peacock, et al. Potential for estimation of body condition scores in dairy cattle from digital images. *Journal of dairy science*, 91(9):3439–3453, 2008.

[6] S. Chetlur, C. Woolley, et al. cudnn: Efficient primitives for deep learning. *CoRR*, abs/1410.0759, 2014.

[7] L. Deng, D. Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.

[8] J. Ferguson, G. Azzaro, et al. Body condition assessment using digital images. *Journal of dairy science*, 89(10):3833–3841, 2006.

[9] J. D. Ferguson, D. T. Galligan, et al. Principal descriptors of body condition score in holstein cows. *Journal of Dairy Science*, 77(9):2695–2703, 1994.

[10] A. Fischer, T. Luginbühl, et al. Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in holstein dairy cows. *Journal of dairy science*, 98(7):4465–4476, 2015.

[11] I. Halachmi, M. Klopčič, et al. Automatic assessment of dairy cattle body condition score using thermal imaging. *Computers and electronics in agriculture*, 99:35–40, 2013.

[12] M. Hansen, M. Smith, et al. Non-intrusive automated measurement of dairy cow body condition using 3d video. In *Proceedings of the Machine Vision of Animals and their Behaviour (MVAB)*, pp. 1.1–1.8. BMVA Press, 2015.

[13] A. Heinrichs, C. Jones, et al. Body condition scoring as a tool for dairy herd management. Tech. rep., Penn State College of Agricultural Sciences.

[14] S. Hijazi, R. Kumar, et al. Using convolutional neural networks for image recognition, 2015.

[15] W. Kellogg. Body condition scoring with dairy cattle. Tech. rep., University of Arkansas, Division of Agriculture.

[16] A. Krizhevsky, I. Sutskever, et al. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105. 2012.

[17] M. Krukowski. *Automatic determination of body condition score of dairy cows from 3D images*. Master's thesis, 2009.

[18] O. Markusfeld, N. Galon, et al. Body condition score, health, yield and fertility in diary cows. *The Veterinary Record*, 141(3):67–72, 1997.

[19] J. R. Roche, N. C. Friggens, et al. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *Journal of dairy science*, 92(12):5769–5801, 2009.

[20] J. Salau, J. Haas, et al. Feasibility of automated body trait determination using the sr4k time-of-flight camera in cow barns. *Springer Plus*, 3:225, 2014.

[21] A. N. Shelley. *Incorporating machine vision in precision dairy farming technologies*. Ph.D. thesis, University of Kentucky, 2016.

[22] R. Spoliansky, Y. Edan, et al. Development of automatic body condition scoring using a low-cost 3-dimensional kinect camera. *Journal of Dairy Science*, 99(9):7714–7725, 2016.

[23] C. Szegedy, W. Liu, et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9. 2015.

[24] E. Wildman, G. Jones, et al. A dairy cow body condition scoring system and its relationship to selected production characteristics. *Journal of Dairy Science*, 65(3):495–501, 1982.

Computer vision based system for apple detection in crops

Mercedes Marzoa Tanco[1], Gonzalo Tejera[2], J. Matas Di Martino[3]

[1] Universidad de la Repblica
Montevideo, Uruguay
mmarzoa@fing.edu.uy

[2] Universidad de la Repblica
Montevideo, Uruguay
gtejera@fing.edu.uy

[3] Universidad de la Repblica
Montevideo, Uruguay
matiasdm@fing.edu.uy

1 Introduction

Currently there is an increasing need to obtain higher quality products at a lesser cost, thus increasing competitiveness. Developing automatic systems to enable the use of human resources more efficiently in terms of precision, repeatability or time consumed is a good alternative. Also, estimating crops yield helps producers improve the quality of their fruit and reduce operation costs. Managers can use estimation results to plan the optimal capacity for packaging and storage [8]. As an example, the cost of harvesting citrus fully by hand may range from 25% to 33% of the total cost of production [3]. The overall objective of this article is to introduce an automated computer vision system for the detection and counting of red apples in trees. The present work is part of the design of a more complex system for the automatic estimation of crops yield and harvesting. In order to detect pixels belonging to apples, three techniques are evaluated: Support Vector Machines (SVM), K-Nearest Neighbors (k-NN) and a very simple decision tree (DT). As an improvement to the outcome of pixel detection, morphology operations have been used. In order to detect the apples themselves, and given their shape, the Hough transform for circles has been used. Finally, a post processing of the circles found is made, to rule out false positives detections.

2 Proposed Solution

2.1 Construction of the database

A database with 266 high resolution images was created and made publicly available. The database is formed by images acquired using natural light, at “Las Brujas - INIA” experimental station located in Canelones, Uruguay. Each image has an associated binary image where it is indicated, in white, whether the pixel belongs to an apple, or in black, if not. In its turn, there is a file per image with the coordinates of each apple’s center. The data base is available for future works [1].

2.2 Classification of pixels in apple or background

To describe those pixels that belong to an apple and those that belong to the background we use three basic features: Hue, Saturation, and a simple Texture descriptor. We tested three different algorithms: a very simple decision tree, K-Nearest Neighbor and Support Vector Machine[2].

2.3 Counting of apples

After having classified the pixels of an input image as part of an apple or the background, we have a groups of pixels labeled as apples, not their quantity or location. The next step is then to identify the apples within those groups of pixels. In order to do that, we resort to techniques applied to a binary image where the background is set as black and pixels that belong to an apple are set as white. The process we apply consists of the following three stages:(i) Pixels Mask Pre-processing, (ii) Detection of circles, and (iii) Circles Validation.

3 Results

The performance of different methods is evaluated according to the F-measure obtained due to the imbalance nature of this problem. The most accurate algorithm in terms of the F-measure was the K Nearest Neighbor method with approximately 64% of F-measure. Using the mask of pixels classified by decision tree method, morphological operations are applied to achieve the final goal of detecting the apples present in the test images. Part of the images of the data set were used to train and fit the parameters of the solution implemented, while other independent images (never used during the training step) are used for testing the proposed solution. As in the pixel classification step, the final performance of the proposed solution is analyzed in therms of the F-measure. It is important to point out that while the definition of F-measure, Recall and Precision is unique, the meaning of this quantities is different in this second step. In the pixel classification step, we define the TP (true positive) quantity as the number of pixels classified as part of an apple that truly belong to an apple, in the final step, TP is defined as the number of apples correctly detected, while FP is the number of apples detected that are not present in the image (false detections) and FN corresponds to the number of apples present in the image that where not detected by the method. Table 1 summarizes the final results obtained for the detection of apples. In this table, the Recall represents the percentage of apples present in the input image that are correctly detected, and the Precision indicates the percentage of detections give by the algorithm that actually correspond to an apple.

Is important to highlight some difficulties that arises when comparing different apple detection approaches. In first place, different solutions proposed in the literature make use of significantly different setups. For example: some works make uses of stereo pairs of cameras plus high precision positioning systems [11], tunnel like structures [4] to control illumination conditions, hyperspectral cameras [9], or thermal imaging devices [10]. Secondly, the conditions of crop yield also has a great impact on systems performance, for instance, special fruit thinning may significantly simplify the problem. Thirdly, the success measurements uses also present significant variations, for example, in [11] the focus is on the over all apple count, hence false positive detections may be compensated with false negative detections (while the f-measure penalize both).

Method	**Recall**	**Precision**	**F-measure**
[12]	67,9%	100,0%	80.1%
[7]	72,8%	97,2%	83.3%
[5]	46,4%	100,0%	63,4%
ours	92,0%	90,3%	91.5%

Table 1: Recall, Precision and F-measure for different apple detection strategies.

4 Conclusions

This article presents a simple pipeline for the detection of apples in crops using pattern recognition and computer vision tools. The set up consists of a single RGB camera that captures under unconstrained natural illumination conditions in unthinned apple crops.

The detection of apples was made in two big stages. First the classification of pixels and then the detection of apples within the previously classified pixels. Three techniques were studied for the classification of pixels: decision tree, KNN and SVM. The best results were obtained with KNN algorithm, while the decision tree probes to be a very adequate alternative if computational cost or time are very limited. The determinant features to make the classification were tonality, saturation and edge density. As an improvement of the recognition of pixels, morphological operations were used. Once the pixels had been classified, we proceeded to the detection of the apples by using the Hough transform for circles. Finally, the quantity of pixels within the circles found was analyzed to validate the circles detected and significantly reduce the number of false positive.

The main contributions of this work are: The use of robust machine learning techniques by facing the problem as a pattern recognition imbalanced problem. We present an updated review of the current state of the art and create a database with 266 high resolution images which was made publicly available.

There are some evident path in which this work can be pushed forward. For example: the output of multiple classifiers can be combined to improve the over all performance [6]. With the increase of the number of publicly available databases, the design of complex modern classifiers such as deep neural networks will be possible. And finally, instead of processing individual images, sequence of images (video data) can be analyzed ensemble exploiting the temporal correlation of the data.

References

[1] Apple data base - Marzoa M and Caggiano S. https://gitlab.fing.edu.uy/mmarzoa/apple_database.

[2] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] Edward R Dougherty and Roberto A Lotufo. *Hands-on morphological image processing*, volume 71. SPIE press Bellingham, 2003.

[4] A Gongal, A Silwal, S Amatya, M Karkee, Q Zhang, and K Lewis. Apple crop-load estimation with over-the-row machine vision system. *Computers and Electronics in Agriculture*, 120:26–35, 2016.

[5] Wei Ji, Dean Zhao, Fengyi Cheng, Bo Xu, Ying Zhang, and Jinjing Wang. Automatic recognition vision system guided for apple harvesting robot. *Computers & Electrical Engineering*, 38(5):1186–1195, 2012.

[6] Ludmila I Kuncheva. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

[7] Raphael Linker, Oded Cohen, and Amos Naor. Determination of the number of green apples in RGB images recorded in orchards. *Computers and Electronics in Agriculture*, 81:45–57, 2012.

[8] Terence Robinson. The evolution towards more competitive apple orchard systems in the usa. *Department of Horticultural Sciences, Cornell University, Geneva*, 2013.

[9] Omri Safren, Victor Alchanatis, Viacheslav Ostrovsky, and Ofer Levi. Detection of green apples in hyperspectral images of apple-tree foliage using machine vision. *Transactions of the American Society of Agricultural and Biological Engineers*, 50(6):2303–2313, 2007.

[10] Denis Stajnko, Miran Lakota, and Marko Hočevar. Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. *Computers and Electronics in Agriculture*, 42(1):31–42, 2004.

[11] Qi Wang, Stephen Nuske, Marcel Bergerman, and Sanjiv Singh. Automated crop yield estimation for apple orchards. In *Experimental Robotics*, pages 745–758, 2013.

[12] Jun Zhao, Joel Tow, and Jayantha Katupitiya. On-tree fruit recognition using texture properties and color data. In *Intelligent Robots and Systems, 2005.*, pages 263–268, 2005.

Barley recognition under different fertilization treatments using machine learning and UAV imagery data

Sara V. Rodriguez [1], Hugo Jair Escalante[2], Manuel Jimenez[1], Rigoberto Vazquez[1]

[1] Universidad Autonoma de Nuevo Leon,

[2] Instituto Nacional de Astrofísica Óptica y Electrónica,

sara.rodriguezsn@uanl.edu.mx

Abstract

We describe a methodology for barley recognition under different fertilization treatments by using machine learning and images collected from an unmanned aerial vehicle, showing the potential to integrate imagery processing tools with UAV technology for increasing accuracy and reducing timing, and cost in agriculture studies. The study focuses on prediction of fertilization treatment and barley variety, results are encouraging and motivate further research on this topic.

1 Introduction

Food security is one of the biggest concerns of the world. Today, 1 person in 9 is hungry. According with estimates of the Food and Agriculture Organization of the United Nations (FAO), by 2050, the world's population will reach 9.6 billion people, which is 35 percent higher than now. This growing population, in addition to the spread of prosperity across some crowded countries, especially in China and India who are demanding more quantity and quality foods, and the extensive use of biofuels, turns out that feeding the world be a great challenge. In this scenario, agriculture is playing a key role to increase food security index. The agriculture is important not only because provides food but also is the main source of raw materials to major industries. Studies to improve crop yields are therefore essential to meet the increasing pressure for global food demands.

Profitable barley production requires efficient nitrogen fertilization management. N fertilizer applications should be apply to the right time and to the right dosis, but N rates vary for type of seeds, soil and season (temperatures and rainfall) conditions.

The objective of this study is to develop method for barley recognition under different fertilization treatments on field, and thus, identify the best fertilization practices according to the type of barley, and fertilization doses. This research offers the opportunity to automate the collection, processing, and analysis for UAV imagery data through the use of machine learning, saving time, money, and in some cases increasing the accuracy of measurements. We target the recognition of fertilization treatment, and barley variety from aerial images. Our methodology is based on deep learning for feature extraction and standard classification techniques for recognition. We show encouraging results using this methodology that does not relies on any indicator or additional information (e.g., NDVI) for recognition: the deep network learns discriminative features directly from raw pixels.

2 Methodology

2.1 Experimental design

The experiment was conducted at the Agronomy Research Field on the Agronomy Campus of the Universidad Autonoma de Nuevo Leon, located in Marin, Mexico. The barley is the agri product studied. To create various crop growth scenarios, 6 barley (Hordeum vulgare) crops were grown (V1= Cuahutemoc, V2= Menonita, V3= Mezcalera V4= Marín, V5= Chichimeca, and V6= UANL-138), under 3 nitrogen fertilization treatments (0, 100, 200 kg N/ha). 72 rectangular plots were established for harvesting, each one belongs to one type of barley and one level of nitrogen fertilization. The fieldwork consisted of soil preparation with tracking and furrowing operations, using a randomized block design, six furrows of 5 m long and 0.7 m between furrows.

2.2 Experimental platform

Hyperspectral as well as RGB imagery was obtained with a hexacopter with six brushless motors of 700 KVAs and 4 electronic speed controllers. The autopilot that has been used is the Pixhawk of 3D Robotics and in this autopilot has been connected to a DX7s 7-Ch DSMX Radio System of Spektrum and in the same autopilot has been connected a telemetry in order to know in real time all the movements and trajectories of the hexacopter. The hyperspectral sensor selected for this experiment was the Parrot Sequoia which possess five channels, 4 of them collect the following bands: Red, Green, Red-Edge and Near-Infra-Red with 660, 550, 735 and 790 nanometers of wavelengths, all of them with a resolution of 1.2 Mpx. Additionally a RBG channel of 16 Mpx, Figure 1.

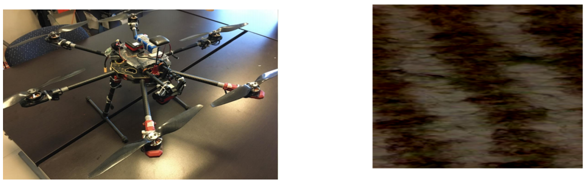


Figure 1. Left: Integrated Aerial Platform. Right: sample image for a particular block.

2.3 Deep-learning based feature extraction from images and classifiers

The image obtained from the flight realized on March 28 of 2017 was used to implement the method for barley recognition. Pre-trained deep convolutional networks (CNNs) were used as feature extractors by adopting a transfer learning methodology. CNNs are neural networks that incorporate convolutional and pooling layers in addition to activation functions. They have been successfully used in a number of applications and they rule the computer vision field nowadays. We used two widely used pre-trained architectures (VGG-19 and VGG) and used them for feature extraction. This process consisted on passing the cropped images taken from the UAV (each image covering a single field), then we used the response from the pre-last layer (RELU) of the CNNs. This resulted in a 4096 dimensional vector that is used as the representation for the corresponding image. Feature vectors are then feed into standard classifiers under the one-vs-one formulation. Figure 2 shows the activation of filters in the CNN when a crop image is feed to the network..

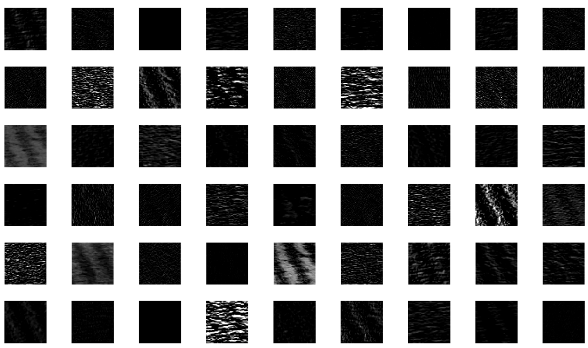


Figure 2. Response for the image in Figure 1 of the first layer of convolutional filters of the CNN.

3 Results and conclusions

We evaluated the performance of the proposed methodology for predicting the variety of barley (6 categories) and the fertilization treatment (3 classes) using the data set of 72 fields-images. A leave one out formulation was adopted for evaluation: training in 71 images and testing in 1 sample, updating the test sample, repeating this process 72 times. We report average recognition performance for different classifiers in Table 1.

Classifier / Problem	Variety	Fertilization
Linear SVM	**38.88 %**	76.38 %
Neural network	37.50 %	**80.55 %**
Naive Bayes	20.83 %	40.27 %
KNN	26.38 %	61.11 %
Random baseline	16.67%	33.33 %

Table 1. Recognition performance for different classifiers using the VGG-based features.

Results obtained with SVM and a neural network are promising for the recognition of variety and fertilization recognition from pixels, respectively. Results are encouraging as more sophisticated preprocessing (e.g., subsampling, image enhancement) feature extraction (e.g., fine tuning) and classification (e.g., ensemble learning) procedures may be used. Additionally, information from pixels can be combined with domain knowledge information (e.g., NDVI) for improving performance. The next step of this research is to address the barley yield prediction problem from pixels. We foresee this is a feasible and very promising field of research.

References

[1] Hunt, E., W. D. Hively, C. S. Daughtry, G. W. McCarty, S. J. Fujikawa, T. Ng, M. T chitella, D. S. Linden, and D. W. Yoel. Remote sensing of crop leaf area index using unmanned airborne vehicles, *Proceedings of the Pecora 17 Symposium*, Denver, CO, 2008.

[2] Ipate, G., Voicu, G., Dinu, I., Research on The Use of Drones in Precision Agriculture U.P.B. Sci. Bull., Series D, Vol. 77, Iss. 4, 2015.

[3] Odido, D., and D. Madara, Emerging Technologies: Use of Unmanned Aerial Systems in the Realisation of Vision 2030 Goals in the Counties, *International Journal of Applied*, v. 3, 2013.

[4] Ponti, M. P., Segmentation of low-cost remote sensing images combining vegetation indices and mean shift, *Geoscience and Remote Sensing Letters*, IEEE, v. 10, p. 67-70, 2013.

[5] Naiqian, Z., Maohua., W., Ning., W., Precision agriculture-a worldwide overview. Computers and Electronics in Agriculture. 113-132-36 (2-3), 2002.

[6] Tripicchio, P., M. Satler, G. Dabisias, E. Ruffaldi, and C. A. Avizzano. Towards Smart Farming and Sustainable Agriculture with Drones: Intelligent Environments (IE), 2015 International Conference on, p. 140-143, 2015.

[7] Uno, Y., Prasher, S.O., Lacroix, R., Goel, P.K., Karimi, Y., Viau, A., Patel, R.M., Artificial neural networks to predict corn yield from Compact Airbone Spectrographic Imager data. *Computers and electronics in agriculture,* 149:161-47, 2005.

Dynamic diet formulation responsive to price changes: a feed mill perspective

Adela Pagès-Bernaus[1], Lluís M. Plà[1], Jordi Mateo[1], Francesc Solsona[1]

Universitat de Lleida
Campus Cappont, 25001, Lleida, Spain
adela.pages@matematica.udl.cat, lmpla@matematica.udl.cat, jmateo@diei.udl.cat, francesc@diei.udl.cat

Abstract

Feed producers offer a large catalog of products highly specialized for different species of animals in order to satisfy customers demand. In a very competitive market, the optimization of the composition of each diet becomes crucial for the economical success of the company. A central planning of the raw material purchases is the best way to take into consideration all the diets in addition to other particular requirements. In this work we present a multi-formulation model which reacts to variation in prices and we test the model in realistic benchmark instances.

1 Introduction

Given that animal feed represents the most significant cost in animal production, the dominant criteria to select the ingredients to produce a particular feed is based on cost. Feed producer companies offer different feed compounds with specific nutritional characteristics to address the needs of a particular animal, age and life stage in the best manner. Nutritionists and veterinaries adjust the requirements for each particular diet. However, there are many combinations of ingredients that can be used to produce a particular diet that fulfills the requirements.

Large feed producers need to purchase large amounts of raw material (such as wheat, barley, corn, ...) on a regular basis, so feed mills can operate smoothly. A multi-formulation model finds the optimal amount of raw material needed for a period to produce the required amount of feed that meets all the nutritional requirements. The multi-formulation model is the extension of the well-known diet problem to account for several diets. It also includes other constraints that link the final recipe for each formulation, such as limits on storage capacity. The main goal of this work is to assist purchase managers in the decisions of the tonnages of commodities to buy.

The strategies adopted to stock up raw materials may range from fix contracts with producers to buy in several type of markets [2]. To take advantage of the cost variations of the raw material in this work we present an extended multi-formulation model. Price for the raw materials is collected from several market places and exchanges (such as the Chicago Board of Trade and other local exchanges). As a first approach, the attributes of the ingredients are assumed to be known based on official tables.

2 Multi-formulation model

A diet consists in finding the combination of raw material that satisfies some nutritional considerations [1]. A cost-minimization solution is usually modeled as a Linear Programming model and it is commonly used by farmers and feed producers. For a single diet, the decision variables are the proportion of each ingredient used to produce a certain quantity of the diet. The set of constraints that limits the composition are:

- The minimum and maximum levels of specific nutrients that the diet needs to satisfy. Examples are a minimum percent of net energy, fiber or protein.
- The maximum content of each possible ingredient.
- Limits on the proportion of ingredients used in the formula. This requirement is imposed to assure that animals will accept the changes that may incorporate the new formulation.

- Other operational constraints related to the minimum amount of raw material to be bought (if used), or storage capacity limits.

Large feed producer companies usually offer feed to different type of animals (such as pigs, ruminants or poultry) and within each type of animal the nutritional needs change with the age and reproductive stage. Overall a feed company may end up producing several dozens of diets. The joint solution of the multi-formulation model results in a large optimization problem.

Moreover, many companies are structured as supply chains. For instance, the pig supply chain (PSC) includes organizations in charge of procurement, production, slaughtering, processing, distribution and marketing of pig meat, derived and by-products to the final consumers [4]. Different PSC agents work together in a coordinated way for the realization of dependent processes leading to pig production and marketing. Feed mill companies are a key agent of this chain structure. While open production and contracting by farmers characterized pig market in the past, nowadays, within a PSC context, this is almost nonexistent. Vertical integration and coordination around feed mills is usually developed by cooperatives and private companies, also called integrators [3]. In these sense, the production planning of a feed mill is crucial for the regular operation of the whole system.

In this work we propose a model to design a strategy to buy the most used raw materials for a future planning period such as the following month. The costs are influenced by the prices of the future prices, logistic costs (supply and demand operates globally) and fluctuations on the exchange rate. All these characteristics are considered in the enriched multi-formulation model. The model is tested in realistic benchmark instances.

References

[1] F. Dubeau, P. O. Julien, and C. Pomar. Formulating diets for growing pigs: Economic and environmental considerations. *Annals of Operations Research*, 190(1):239–269, 2011.

[2] N. Merener, R. Moyano, N. E. Stier-Moses, and P. Watfi. Optimal trading and shipping of agricultural commodities. *Journal of the Operational Research Society*, 67(1):114–126, 2016.

[3] C. Perez, R. de Castro, and M. Font i Furnols. The pork industry: a supply chain perspective. *Brithish Food Journal*, 111(3):257–274, 2009.

[4] S. Rodriguez, L.M. Plà, and J. Faulin. Piglet production inside the Pork Supply Chain Management. A linear optimization model in sow farms. *Annals of Operations Research*, 219(1):5–23, 2014.

Mathematical modeling under uncertainty for the sugar supply chain management in Cuba

Esteban López[1], Lluis Miquel Plà[2]

[1] University of Holguin
Mechanical Engineering Department, University of Holguin, Av. XX Aniversario s/n, 80100 Holguín, Cuba
elopez@uho.edu.cu

[2] University of Lleida
Mathematics Department, University of Lleida, Jaume II, 73, 25001 Lleida, Spain
lmpla@matematica.udl.cat

Abstract

Numerous variables are involved in the sugar supply chain from the fields to the mill: corresponding to the processes of cutting, loading and transporting the cane. This paper proposes a two stage stochastic version of a mixed linear programming model published by the same authors. Uncertainty in weather and harvesting conditions is represented through different scenarios. This model allows to minimize the cost of the transportation, and elaborates a daily schedule of resources allocation like cutting and transportation of sugar cane under Cuban conditions.

1 Introduction

The management of the sugar cane supply chain, and in particular the sugar cane harvest, is a complex logistical operation that involves the cutting and loading of cane at the fields, the transportation by truck or train to the sugar mills, and the unloading of the cane to be processed in the mill [2, 5, 6].

The Cuban sugar industry is characterized by sugar mills able to take supplies of cane from surroundings farms. Sugar cane must be cut when it is ripe, if not sugar cane quality deteriorates. Depending on the quota and available resources in a particular day, the scheduling is proposed by sugar mill managers based on their own expertise. Taking into account daily changes in the amount of cane in the fields, the cane ripeness, the unforeseen failures in machinery, and the performance of harvesters, managers must adapt their schedules daily [3].

As sugar cane is harvested it is transported to the sugar mill. Generally, sugar cane can be conveyed in two different ways: by "direct transportation" with automotive transportation equipment and by the "combined transportation" (Figure 1). The combined transportation uses road transportation means to transport the cane to the collection centers, where the cane is cleaned out of straw, then it is placed in the containers of the railroad, to be carried to the yard of the sugar mill where it waits until it is processed.

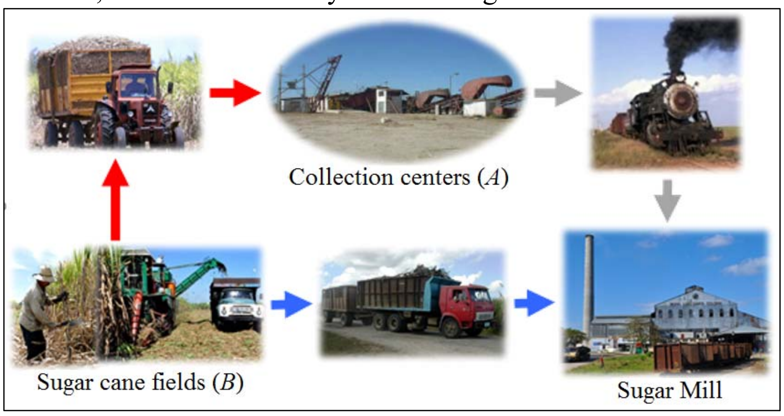


Fig.1. Supply chain of sugar cane to the Sugar Mill in Cuba.

The transportation system has to maintain a constant flow of ripe cane to the mill [1]. The rail system operates 24 hours a day, whereas the harvest period may comprise only a part of the day [4]. Therefore, when at night road transport stops working at night, the rail system assumes all the supply. In this way, the rail system acts as a storage room for cut cane, allowing the creation of a reserve that satisfies the demand of the sugar mill, while road transportation is covering other routes at the same time or when it has stopped at night. Unless a mill failure or break down occurs, railway transportation allows the sugar mill to work 24 hours a day without interruption; however, every 10 days the sugar mill is stopped for technical maintenance.

2 Stochastic approach to the problem

The deterministic model was formulated by [2] as a mixed-integer linear programming model. This paper proposes a stochastic extension of this model. The main risk affecting sugar cane harvesting and subsequent production is the rain. Rain make sugar cane wet and heavier and cutting and transportation to the mill slower. For simplicity, three possible scenarios are considered: no rain ($S1$ is the probability for no rain), little rain ($S2$ is the probability for little rain) and moderate rain ($S3$ is the probability for moderate rain).

The model is formulated for a working day. Here, the decision variables are represented by $X_{i,j,k,l,s}$, the subscripts i, j, k and l have the same means as in the deterministic problem. A new subscript is added: the s subscript ($s = \{1, 2, 3\}$). It represents the possible scenarios: $s = 1$ for no rain scenario, $s = 2$ for little rain scenario and $s = 3$ for moderate rain scenario.

The decision variables have a combinatorial nature, and not every combination is possible; to define those that will be feasible in the model formulation, some rules are necessary:

- The variables determining routes (both for road and rail transportation) where an origin is also the destination are not considered;
- In case the origin is a storage facility ($i = 1$ to A), the only destination admitted is the sugar mill ($j = 1$). The storage facilities will not transfer cane between them, and only unloading it in the swing bolster is allowed;
- The sugar cane fields as origins will admit any destination ($j = 1$ to $j = A + 1$);
- The variables presuming the railway transportation ($k = 1$) will only be defined for the combination with the sugar mill ($j = 1$), the subscripts $l = 1$ and $s = 4$.

2.1 Constraints

Main constraints refer to constraints always present in the different formulations of the problem. The core of the problem can be solved for one working day, and ignoring the schedule hour by hour. These are constraints including only continuous variables. The constraints of the mathematical model are classified in the following groups:

- Supply of cane to the sugar mill for a working day;
- Capacity of the collection centers;
- Conservation of flow-through storage facilities;
- Capacity of transportation by road transportation means;
- Production of the sugar cane fields;
- Cutting capacity of different teams.

2.2 Objective function in the stochastic approach

The objective is the minimization of daily transportation cost. The economic coefficients ($C_{i,j,k,l,s}$) of the objective function establish the transportation cost of sugar cane, related to the distances and the transportation means used in each case and the possible scenarios.

Quality aspects are considered by means of an opportunity coefficient ($0 < Co_{i,s} \leq 1$) determined empirically by the decision-maker, and by establishing minimum quantities of cane processed just in time to preserve cane freshness. It represents the preference to cut a sugar cane field i. By default, it is assumed that all fields susceptible of harvest have a similar maturation level and $Co_{i,s} = 1$.

$$Min\,\mathrm{C} = \sum_{i=1}^{A} C_{i,1,1,1,4} \cdot X_{i,1,1,1,4} + \sum_{i=A+1}^{A+B} \sum_{j=2}^{A+1} \sum_{k=2}^{K} \sum_{l=2}^{L+C+1} (S1 \cdot (C_{i,j,k,l,1} \cdot Co_{i,1} \cdot X_{i,j,k,l,1}) + S2 \cdot (C_{i,j,k,l,2} \cdot Co_{i,2} \cdot X_{i,j,k,l,2})$$

$$+ S3 \cdot (C_{i,j,k,l,3} \cdot Co_{i,3} \cdot X_{i,j,k,l,3})) \tag{12}$$

2.3 Total constraints and variables

2.3.1 Constraints

Daily mill supply:	2
Capacity of storage facilities:	A
Cane coming in and leaving collection centers:	A
Transportation means:	$(K - 1)$
Production of the sugar cane fields:	B
Constraints of cutting means:	$(L + C) \cdot B$
Total:	$2 \cdot A + K + B \cdot (1 + L + C)) + 1$

2.3.2 Variables

Continuous variables delivering cane to the mill:

$X_{i,1,1,1,4}$:	A
$X_{i,1,k,l,s}$:	$B \cdot (K - 1) \cdot (L + C) \cdot 3$

Continuous variables delivering cane to the collection centers:

$X_{i,j,k,l,s}$:	$B \cdot A \cdot (K - 1) \cdot (L + C) \cdot (s - 1)$
Total:	$3 \cdot (K - 1) \cdot (L + C) \cdot (B \cdot (1 + A)) + A$

3 Conclusions

The proposed model integrates rail and road transportation systems emphasizing the reduction of transportation cost under uncertain weather affecting sugar supply chain operation. At the same time, it controls sugar cane freshness through the constraints of minimum supply to the sugar mill with direct transportation.

Uncertainties during the harvesting seasons have to be considered to prevent undesired events impacting on sugar production. Hence, scenarios are built considering the probability of rain since it is one of the main factors affecting negatively sugar production. The weight of the sugar cane increases by the augmentation of water content and furthermore, the speed of transportation decreases if rain occurs.

Current technology permits to solve huge linear programming models, but managers find them difficult to handle. Because of this reason, it is helpful to elaborate stochastic extensions of deterministic models to value the risk of ignoring uncertainties inherent to harvesting.

References

[1] A.J. Higgins. Australian sugar mills optimize harvester rosters to improve production. *Interfaces* 32(3):15–25. 2002.

[2] E. López, Silvia Miquel and Lluis M. Plà. Sugar cane transportation in Cuba, a case study. *European Journal of Operations Research*. 174 pages 374–386. 2006.

[3] E. López and Lluis M. Plà. Chapter V: Optimization of the Supply Chain Management of Sugarcane in Cuba. *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, Springer Science + Business Media, LL, International Series in Operations Research & Management Science 224, DOI 10.1007/978-1-4939-2483-7_5. 2015.

[4] E. López, Silvia Miquel and Lluis M. Plà. El problema del transporte de la caña de azúcar en Cuba. *Investigación Operacional.* Vol. 25, No 2, pages 148 – 157. 2004.

[5] R. Pavia and R. Morabito. An optimisation model for the aggregate production planning of a Brazilian sugar and ethanol milling company. *Annals Operations Reserarch* 161:117–130. 2009.

[6] S.D. Jena and M. Poggi. Harvest planning in the Brazilian sugar cane industry via mixed integer programming. *European Journal of Operation Research* 230:374–384. 2013.

Planning tool for the multisite pig production system based on stochastic optimisation

Esteve Nadal-Roig[1], Lluis M. Plà[2]

[1] Universitat de Lleida
c/ Jaume II, 73, 25003, Lleida (Spain)
enr1@alumnes.udl.cat

[2] Universitat de Lleida
c/ Jaume II, 73, 25003, Lleida (Spain)
lmpla@matematica.udl.cat

Abstract

This paper presents a planning production tool based on a stochastic programming model for helping pig managers to take decisions in a three-site pig production process. The model is intended for practical use and based on a real instance. The main sources of uncertainty considered by the model are related to fertility and litter size rates. Practical results include sow replacement & purchases, transfers of animals between farms, batch management, occupancy rate of facilities, and the management of pig deliveries to the abattoir. Because of its mixed-integer nature we highlight the complexity of the model in terms of computational effort.

1 Introduction

The structure of the pork sector is experiencing changes in the last years [2]. The traditional pig production based on small family farrowing-to-finish farms, where single farmers are in charge of the entire pig production process, is being transformed to a bigger, more controlled and industrialized environment where farmers are part of integrators and/or cooperatives and covering only a part of the production process [7]. At the same time, integrators and cooperatives tend to integrate and coordinate their operations into pork supply chains involving many farms which present competitive advantages and benefits since it helps to reduce the risk and uncertainty and creates value [4]. Hence, activities like planning piglet production, the control of animal's stock over time and the scheduling of transfers among farms deserve attention of chain managers who allocate time and resources to solve those questions properly.

In this context, a modern pig production system evolved from old farrowing-to-finish farms is represented by a three-site multi-farm system with three phases. The first phase focuses on producing the piglets, the second phase focuses on rearing the piglets and the third and last phase focuses on fattening the pigs and delivering them to the abattoir. For each of these phases, a set of specialized farms located separately (i.e. sow farms, rearing farms and fattening farms respectively) are involved. Each one has their own characteristics, facilities and location. Therefore, transportation between phases is mandatory.

[6] stated that sow farms involve the most important and complex activity in the production process due the required control and efficiency of sows determining piglets' production. In this phase, three different facilities are considered: breeding, gestation and lactation. In the breeding facility, sows are inseminated and controlled to confirm pregnancy. If confirmed, sows are transferred to the pregnancy facility. Otherwise, sows remain for re-insemination. Finally, lactation facility is where piglets born and live till weaning. At any time, sows which are not considered productive are sent to the abattoir and replaced by new ones.

Contributions for helping to the decision-making process in the entire pig production systems involving different farms are not extensive [7]. For instance, [5] proposed a first mixed-integer linear programming model to optimize the entire production process following a three-site structure. [1] developed a tool based for helping the decision-making of the chain managers reformulating [5] by including the transportation constraints. Thus, the aim of this work is to extend the functionality of [1] by adding new capabilities considered essential for this industry and uncertainty in the fertility and litter size rates.

2 Characteristics of the model

In our work, the model maximizes the total revenue calculated from the income of sales to the abattoir minus the production costs over the time horizon considered. Production cost depends on purchases, feeding system, veterinary and medical care, labor and transportation. We use the structure and data of a real integrator based in Catalonia (Spain) which farms are grouped by 9 sow farms, 22 rearing farms and 131 fattening farms and one abattoir. Each farm has its initial inventory. Transportation between farms is necessary which is performed by trucks. The integrator subcontracts this activity to a specialized company sending in a weekly basis the schedule of the trips to be done. Trucks' capacities of animals vary depending on each stage of the productive process. Transportation cost is set by Euro/Kilometer.

In sow farms, the sows are inseminated and controlled for 3 weeks. In case of confirmation, the sows are transferred to the pregnancy facility, after 9 farrowing stages are not expected to be inseminated anymore. Otherwise the sows remain for re-insemination which in case of no confirmation after three attempts, the sows are sold to the abattoir. In both cases, sows sold to the abattoir are replaced by new ones via purchasing. Purchases cost are taken into consideration to the standard cost marketplace. Abortion, fertility and litter size rates are parameters given by the integration according historical data. Once the piglets are weaned, they are transferred to the rearing farms to be fed for 6 weeks to ensure their correct development. Finally, in the third phase, piglets are transferred to the fattening farms for a maximum of 18 weeks (stages) at a weekly cost per pig. Fattening farms can be filled in a continuous flow or using the AIAO (batch) management, considered as an industry best-practices for large facilities helping to curb the spread of illness and diseases [3]. The aim of this phase is to sell pigs to the abattoir once they have reached a marketable weight. This means, it is allowed to transport pigs to the abattoir which weight is more than 100 kilograms although they haven't reached the optimal weight. This is from week 15 to 18.

The abattoir represents the pig demand. However, in our case study, the abattoir has enough capacity to slaughter all pigs produced. Sales price is defined by penalties or bonus applied depending on lean content and carcass weight according SEUROP.

The productive process requires decisions in the first week of the time horizon that are constrained by the uncertainty in some of the parameters of the model. Those decisions, called in a stochastic model as first-stage decisions are related to the number of sows replaced & purchases, and the transfers of piglets to be done through the entire production process according the farms' stock and location. Uncertainty is present in the sow fertility and litter size rates that might vary depending of environmental factors, animal welfare, diseases and season. Therefore, those have a direct impact in the overall production [8].

3 Results and conclusion

The tool is able to provide a vision to the chain managers of the flow of the animals during the time horizon specified and a weekly transportation schedule in all the phases of the production process. Also it determines the optimal farms' performance based mainly on the capacity and location. Simulations by adding, removing and changing farms to study the production process helps managers to take decisions in the acquisition of new farms.

In the sow farms, the tool presents a schedule of sow's replacement, the purchase needs over the time horizon and the pigs' production. In the fattening farms, the tool provide to the chain managers a batch management schedule for all the time horizon and the deliveries of pig to the abattoir based on the marketable time window.

All this information including economic indicators (like costs, income and benefit) are also present in a multidimensional data structure based on stages of the model, scenarios, weeks, stage and farms level for allowing managers to have both, a specific and aggregate control of the production process.

In terms of performance, despite of the good results obtained on small instances, like the one presented in this paper, the model fails in terms of execution time when the number of farms of the production processes increases due its integer formulation. At this time manager can solve the model for taking decisions but creates difficulties because of the time spent for executing the model. Heuristics, in order to improve the model performance without affecting the quality are not in the scope of this work but will be considered in the future.

Finally, and as a future work, the presented model can be used as a baseline to adding functionality (new one or already done form other authors) and therefore to extend the pig production process and its supply chain.

References

[1] E. Nadal-Roig, L. M. Plà. Multiperiod planning tool for multisite pig production systems, *Journal of Animal Science,*92 (9), 4154–41, 2014

[2] R. Nijhoff-Savvaki, J.H. Trienekens, S.W.F. Omta. Drivers for innovation in niche pork netchains: a study of UK, Greece and Spain. *British Food Journal*, 114: 1106-1127, 2012

[3] J. Ohlmann, P. Jones. An integer programming model for optimal pork marketing. *Annals of Operations Research*, 190:271–287, 2011

[4] C. Perez, R. de Castro, M. Font i Furnols. The pork industry: a supply chain perspective. *British Food Journal*, 111(3), 257-274, 2009

[5] L.M. Plà, Romero. Planning modern intensive livestock production: The case of the Spanish pig sector, *International Workshop in OR*. La Havana. Cuba

[6] L.M. Plà, J. Faulín, S.V. Rodríguez. A Linear programming formulation of a semi-Markov model to design pig facilities. Journal of the Operational Research Society, 60(5), 619-625, 2009

[7] S.V. Rodríguez, J. Faulin, L.M. Plà. New opportunities of Operations Research to improve pork supply chain efficiency, *Annals of Operations Research,* 5-23, 2012

[8] S. V. Rodríguez, V. M. Albornoz, L. M. Plà. A two-stage stochastic programming model for scheduling replacements in sow farms, *TOP* 17 (1) 171-189, 2009

Analysis of decomposition parameters of green manure in the Brazilian Northeast with Association Rules Networks

Dario Brito Calçada[1], Solange Oliveira Rezende[1] Mauro Sergio Teodoro[2]

[1] Universidade de São Paulo
São Carlos - SP - Brazil
dariobcalcada@usp.br
solange@icmc.usp.br

[2] EMBRAPA Meio Norte
Parnaíba - PI - Brazil
mauro-sergio.teodoro@embrapa.br

Abstract

Modern agricultural processes are increasingly looking at the use of chemicals, so the constant search for organic alternatives to fertilization becomes frequent. The use of data mining using association rule networks (ARN) can aid in the analysis of the parameters involved in choosing which plant to use as green manure. In this work, an analysis of the parameters of green manures used in the Brazilian Northeast is presented, demonstrating the applicability of the computational technique as well as its use to gain in productivity.

1 Introduction

Inadequate processes of occupation of arable areas and the need for rapid food production, coupled with economic interests in the pursuit of profitability in the agricultural sector, have contributed to the worsening of environmental degradation causing severe changes in the physical, chemical and biological attributes of the soil causing a drop in the productive potential [3].

It is possible to practice organic agriculture only to substitute inputs used in conventional agriculture, however, in ecologically based agriculture the sustainable production is possible, in which the aim is to intensify the free natural functions of the ecosystem [2].

With the problem of ecological production versus productivity, one must understand the importance of using plant species, known as green fertilizers, capable of attributing improvements to the production environment, since chemical inputs are not allowed in ecologically based agriculture [8].

Leguminosae are extremely important as green manure, but the greatest difficulty encountered for the use of these species is related to the time of decomposition of this type of plant, which directly affects the productivity of the crop. The choice of the plant type is related to the desired degradation time [6].

This work proposes the use of extraction of patterns for the discovery of parameters directly related to the half-life rate of legumes used as green manure in the Brazilian Northeast.

Data mining techniques, in particular, mining association rules may contribute to the study of parameters related to agroecological production [5]. The discovery of association rules is a data mining technique that seeks to identify certain patterns of data in large databases, allowing, after their interpretation, to acquire specific knowledge about the problem under analysis [7].

An association rule characterizes how much the presence of a set of elements in the records of a database implies in the presence of some other distinct set of elements in the same records [1]. The format of an association rule can be represented as an implication LHS $\Rightarrow$ RHS, where LHS and RHS are, respectively, the Left Hand Side and Right Hand Side of the rule, defined to disjoint sets of items.

For each rule (LHS $\Rightarrow$ RHS), extracted from a set of transactions T, a support value (*sup*) is given that checks the strength of the association LHS and RHS in relation to the total items. The confidence values (*conf*) measures the strength of the logical implication of the rule.

1.1 Association Rules Network (ARN)

Proposed by [9], the central idea of ARNs is that the association rules discovered by the mining algorithm can be synthesized, pruned, and integrated in the context of specific research objectives. In particular, if there is a variable of interest ("target" or "objective"), a network can be formed with the most relevant variables related to the objective and, afterwards, to elaborate a structure that can be tested using statistical methods, i.e., to couple a data mining task with statistical analysis.

As described by [4], ARNs use as a representation a *B-graph* (backward-directed hypergraph), which after the pruning processes, can transform the ARN according to the objective.

2 Decomposition of Green Manure and ARN

The work was conducted during the second semester of 2015, at Embrapa Meio-Norte/UEP Parnaíba, (0305'S, 4146'W and 46.8m altitude).

Seven types of legumes were planted: *Crotalária breviflora*, *Crotalária juncea*, *Crotalária mucronata*, *Canavalia ensiformis L.*, *Cajanus cajan* Fava Larga, *Cajanus cajan* IAPAR 43 e *Tephrosia candida*.

At 120 days, plant height (AP) parameters were determined; Fresh shoot mass (MFPA); Dry shoot mass (MSPA); Fresh root mass (MFR) and Dry root mass (MSR). Germination (G), flowering (F) and pod formation (PV), as well as the collector diameter (DC) and number of branches per plant (NR) were also evaluated.

The residual decomposition constant (k) was calculated for each species, following the simple exponential model used by [10], as well as the half-life for the decomposition evaluation, expressing the time period, in days , required for half of the material to decompose.

After the calculations, all parameters were categorized from 1-6 (one to six) according to the values obtained in the experiments, and then the mining of the association rules was performed with values of *minsup* = 0.3 and *minconf* = 0.5, since these measures were the ones that presented a better number of rules. With the generated rules, the construction of the respective ARN was made.

3 Results and Discussion

The Association Rule Network target was the "HalfLife=6.0" (Figure 1), which indicates a longer half-life, resulting in a longer decomposition time of the green manure.

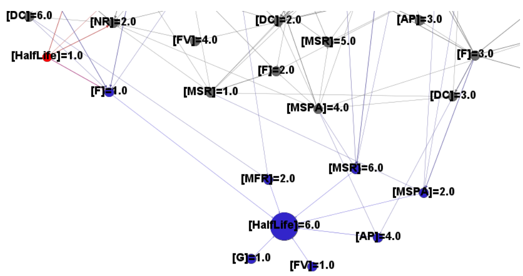


Figure 1: Association Rule Network clipping with target "HalfLife = 6.0"

By verifying the nodes with level 1 (one), i.e. directly connected to the target, one perceives 7 (seven) proper conditions for a greater time of decomposition. First stands out nodes without predecessors, "[G]=1.0" and "[FV]=1.0". It can be inferred that plants with a germination time and a formation of pods in a shorter period tend to decompose more slowly and thus are important characteristics for the evaluation of new compounds.

The fresh root parameter (MFR) presents the "[MFR]=2.0" node, indicating a low rate for this index in plants with longer half-life. In relation to the "[MSR]=6.0" node, it is also inferred its connection with high values for all other mass items (MFR, MSR and MSPA), which corroborates to the search for species that promote a high index of mass in their root and air compositions.

The nodes “[MSPA]=2.0” and “[F]=1.0” undergo lower half-life influences, “[HalLife]=3.0” and “[HalLife]=1.0”, respectively, leading to the need for further study. A decomposition rate in category 4.0 (four) was also observed for the plant height parameter (AP).

4 Conclusion

With the mining through the use of ARN, it was possible to generate the discovery of a knowledge directly linked to studies of green fertilizers, as well as positively influence the choice of the plant according to the crop, and therefore boosting productivity.

For future work, Mining will be performed with other types of plants that can be used as green manure, and compare it with the productivity of the crop in which each species is commonly handled.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. *Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of Twentieth International Conference on Very Large Data Bases, VLDB*, pages 487–499, 1994.

[2] E. J. Ambrosano, F. Rossi, N. Guirado, E. A. Schammass, T. Muraoka, P. C. O. Trivelin, and G. M. B. Ambrosano. Adubação verde na agricultura orgânica. In *Filho, O. F. de L; Ambrosano, E. J.; Rossi, F.; Carlos, J. A. D. Adubação Verde e plantas de cobertura no Brasil: fundamentos e prática.*, chapter 15, pages 45–80. EMBRAPA, Brasília, 1 edition, 2014.

[3] A. Calegari. Perspectivas e estratégias para a sustentabilidade e o aumento da biodiversidade dos sistemas agrícolas com o uso de adubos verdes. In *Filho, O. F. de L; Ambrosano, E. J.; Rossi, F.; Carlos, J. A. D. Adubação Verde e plantas de cobertura no Brasil: fundamentos e prática.*, chapter 1, pages 21–36. EMBRAPA, Brasília, 1 edition, 2014.

[4] S. Chawla. Feature Selection, Association Rules Network and Theory Building. *JMLR: Workshop and Conference Proceedings - The Fourth Workshop on Feature Selection in Data Mining*, 10:14–21, 2010.

[5] F. M. M. de Barros, S. R. de M. Oliveira, and L. H. M. de Oliveira. Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas sobre cana-de-açúcar. *Bragantia*, 72(4):387–395, 2013.

[6] L. F. Garcia. Introdução e avaliação de leguminosas para adubação verde em solos arenosos de Tabuleiros Costeiros do Piauí. *Rev. Fac. Agron. (Maracay)*, 28:93–103, 2002.

[7] T. Le and B. Vo. The lattice-based approaches for mining association rules: a review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(4):140–151, 2016.

[8] C. R. Martins and I. Barros. Intensificação ecológica da fruticultura: sistema de produção ecologicamente intensivo de coco e citros, na Região Norte e Nordeste do Brasil. In *IX Congresso Brasileiro de Agroecologia*, volume 10, pages 1–5, Belém - PA, 2015.

[9] G. Pandey, S. Chawla, S. Poon, B. Arunasalam, and J. G. Davis. Association Rules Network: Definition and Applications Gaurav. *Statistical Analysis and Data Mining*, 1(4):260–179, 2009.

[10] C. de P. Rezende, R. B. Cantarutti, J. M. Braga, J. A. Gomide, J. M. Pereira, E. Ferreira, R. Tarré, R. Macedo, B. J. R. Alves, S. Urquiaga, G. Cadisch, K. E. Giller, and R. M. Boddey. Litter deposition and disappearence in Brachiaria pastures in the Atlantic Forest region of the south of Bahia, Brazil. *Nutrients Cycling in Agroecosystems*, 54:99–112, 1999.

Dealing with derivatives for water quality management

Sira M. Allende Alonso [1] Carlos N. Bouza – Herrera[2], Rajesh Singh[2]
[1]Universidad de La Habana, La Habana, Cuba
Sira@matcom.uh.cu
[21]Universidad de La Habana, La Habana, Cuba
bouza@matcom.uh.cu
[3]Banaras Hindu University, Varanasi, India
rajesh.singh@monash.edu

ABSTRACT

When dealing with data provided by sensors we must deal with Big Data. That is the case in many environment applications. It is often needed to estimate the Average Derivative and the problem is concerned with non parametric regression fitting and the reliability of it is based on the point-wise consistency of an estimator of a probability density function. Our contribution is concerned with modeling the optimization of the estimation of a derivative using suitable models and softwares. We develop a study of the data provided by satellite information to determine the level of eutrophication of a basin

Keywords: Average Derivative, nonparametric density estimation, point-wise convergence

1. Introduction

The challenge of using Big Data is posed by the fact that we must deal with data where scale, diversity, and complexity impose using new architecture, techniques, algorithms, and fresh looks to analytics. Nowadays Big data is of common use in business it is arriving constantly and from multiple sources. Extracting the regularities present in big data needs of using efficient computational means to obtain insights in the patterns, trends, and associations. A look to agriculture evidence the need to improve the available tools for a better handling with the particularities of the different data types, the finding of entities of interest and to develop analyses using existing models. The analysis of the quality of fresh water is very important for the management of water use. For agriculture that is one of the key aspects for managing. An important problem is to evaluate the eutrophication of the water sources as it allows evaluating the quality, nutrition status and organic pollution extent of the water. The common monitoring of Chl-a is to collect samples in the sites. Sampling field water is very costly, affected by measurement errors and time-consuming. Using remote sensing data is possibility for obtaining information on variables that may be used for predicting Chl-a. The reliability of the data is increased, the costs diminished but the information collected from remote sensing, as that provided by satellites, poses a Big data problem to analysts. Analysts want to obtain a classification of the sources. This problem may be modeled using derivatives for predicting Chl-a. The main objective of this paper is analyzing the data studied by Allende et al. (2016) and evaluate how different clustering methods behaves in clustering and which models are better for fitting estimations of the derivatives

The data were divided. A training sample was used for fixing the models and clusters and a validation sample was use for evaluating the different alternatives. The true values of Chl-a were known a them allowed to classify the records into one of the following categories:

A= “is a source of potable water”,

B=“is a source usable only for agriculture and similar proposes”

C=“is source of highly contaminated water”.

The clustering procedures were based on the classic statistical idea of classifying using the minimum distance of the variables to the mean vector of the class . The regression fitting used the kernel of Epanechnikov.

2. The Clustering

Classification into clusters is a well-known multivariate technique. We may consider that classifying is the process of learning from the data. From it we may derive how appropriate is a model for describing the behavior of the phenomena in predetermined classes of data, see Brooks (2010). Once the model is built, it can be used to classify new data. Hence, a goal of Classification is sorting observations into two or more pre-labeled classes. When dealing with Big Data is important deriving a rule to optimally assign a large number of

new objects to the classes. Then each item is identified by its membership. Once the clusters are determined we may find a center for each group of data sets. They are points whose parameter values are the mean of the parameter values of all the points in the clusters. The algorithms output allow obtaining a statistical description of the cluster: centroids and the number of components in each cluster. Algorithms are used to classify new data using the distance between the corresponding point and the clusters centroids. .Each record contains a set of attributes and the collection of records in the training sample is used to model how to assigns the belonging of the record k to a certain class $c_j\ j = 1, \dots, J$. The model must be able to assign unseen records to a class accurately. The test set is used to validate the accuracy of the classification. the model performs well if the classifying of new objects is correct with a high frequency.
We decide $i \in C_j$ if $\underline{x}_i$ is close to a certain predetermined point $\underline{x}_j$ of the cluster C_j. The closeness is measured by a certain distance o similarity measure. In our case we deal with continuous variables then we consider a distance measure.

Take m as the size of the validation sample and $m_{ij} = \begin{cases} 1 \text{ if i is uncorrectly classified in } C_j \\ 0, \text{ otherwise} \end{cases}$

We compute the observed proportion $f_j = \frac{\sum_{i=1}^{m} m_{ij}}{m}$. Then we may estimate the probability of misclassification by: $f = \sum_{j=1}^{J} f_j$. The smaller is f the better is the clustering.

Distances used for clustering

We used the following distances

Euclidian $D_{E:i,j} = \sqrt{\sum_{k=1}^{p} \left(x_{ki} - x_{kj}\right)^2}$; City-Block .$D_{CB:i,j} = \sum_{k=1}^{p} \left|x_{ki} - x_{kj}\right|$; Mahalanobis $D_{M:i,j} = \sqrt{\sum_{k=1}^{p} \frac{\left(x_{ki} - x_{kj}\right)^2}{s_k^2}}$, s_i is the standard deviation; Marczewski – Esteinhaus,:$D_{MS:i,j} = \frac{\sum_{i=1}^{p} |x_{ij} - x_{ik}|}{\sum_{i=1}^{p} MAX\,\{x_{ij}, x_{ik}\}}$; Genetic: $D_{IG:i,j} = \frac{\sum_{i=1}^{p} x_{ij} x_{ik}}{\sqrt{\sum_{i=1}^{p} x_{ij}^2 \sum_{i=1}^{p} x_{ik}^2}}$

3. The study of a basin

The study of the quality of the water in basins is of importance in agriculture studies. A key quality factor is ow high is the concentration of chlorophyll. It is considered as the most important parameter because it may be used for evaluating not only the water quality but the nutrition status and the extent of organic pollution extent. The satellites report to the biologist the hyper spectral reflectance which provides information on the water components as chlorophyll-a. Considering the volume of the data is avoiding the use of spectroscopic analysis and use of the derivative methods. the first-order derivative is able to remove pure water effects and the second-order derivative can remove suspended sediment effects. using the first-order derivative spectrum performs better than the traditional band ratio mode, results support accepting that the first-order derivative model is also better than the single-band and band ratio model. These results suggest that the volume of the data coming for remote sensors may be diminished considerably and use only the reflectance at the wavelengths $\lambda_1, \lambda_2, \lambda_3$ *and* λ_4. In the sequel we will denote them as X_1, X_2, X_3 *and* X_4.
We pose that the relation among the variables may be described as:
Clorophyll concentration at measurement t=$CC_t = c + m(\vec{X}_t) + \varepsilon_t = c + \sum_{b=1}^{4} f_b\ (X_b)$.
Take the Kernel based non parametric regression function

$$\widehat{m_H}(\vec{x}) = \begin{cases} \sum_{t=1}^{T} \frac{CC_t K_H\left(\frac{\vec{x}_t - \vec{x}}{h}\right)}{\sum_{t=1}^{T} K_H\left(\frac{\vec{x}_t - \vec{x}}{h}\right)}, & K_H\left(\frac{\vec{x}_t - \vec{x}}{h}\right) \neq 0 \\ 0 \text{ otherwise} \end{cases}$$

Bouza et al. (2016) recommended for fitting the model using the kernel $K_E(u) = \frac{3}{4}(1 - u^2)I(|u|) \leq 1$, Epanechnikov. Bouza et al. (2016) followed the experimental design used by Cheng et al. (2013) and generated randomly the X_h´s using of 12 035 measurements obtained from an environmental satellite used for

monitoring. They used the model tuning method of Zimba and Gitelson [2006] and the four-band model was developed based on the proposal of Le et al. [2009]. . The main result was determining that the kernel of Epanechnikov had the best performance. A training sample of 6 000 records was selected randomly from the population. The data were classified in C_1, C_2 or C_3 using the real values of CC_t. We use both Epanechikov´s kernel and Yang et al.´s method for predicting CC_t. in the validation sample.
The training sample was classified using the following classification rule

$$C_1 = \{t|CC_t < 40mg.m^3\}, C_2 = \{t|40mg.m^3 \leq CC_t < 67,5\, mg.m^3\}, C_3 = \{t|CC_t \geq 67,5\, mg.m^3\}$$

The validation sample was classified with respect to the mean vectors of the classes using the rule

$$y_{iC} = \begin{cases} 1 \text{ if } D_M(i, C_1) < 4,50 \\ 2 \text{ if } 4,50 \leq D_M(i, C_1) < 15,50 \\ 3 \text{ if } D_M(i, C_1) \geq 15,50 \end{cases}$$

We computed the proportion of wrong classification provided by each minimal set

$$f(M(t,j)) = \frac{\text{number of missclassifictaions of records using Mt in the class j}}{\text{number of record in the sample}}, t = 1, \ldots, 4, j = 1,2,3$$

The estimation of the probabilities of misclassification are given in table 2 for the different distances.

Table 2. Estimated Misclassification Probabilities

		Classes		
		j		
	1	2	3	Total misclassification probability
Distance to the mean f_j	0,119	0,136	**0,043**	0,259
Euclidean	0,472	0,352	**0,027**	0,220
City block	0,555	0,262	**0,061**	0,144
Mahalanobis	0,583	0,363	**0,035**	0,473
Marczewski – Esteinhaus	0,138	**0,102**	**0,100**	**0,120**
Genetic	**0,107**	0,131	0,132	0,411

4. Conclusions

The best overall classification method is using the “Marczewski – Esteinhaus Distance” as well as for the usable for agriculture; for the highly contaminated water the best method is the Euclidean; while the behavior of Genetic provides the smaller misclassification probability when classifying the potable water for agricultural use .

References

1. Allende, S. M., D. C. Chen, C. N. Bouza, J. M. Sautto and A. Santiago (2016): Estimation of derivatives, from economy to environment: a study of the management of eutrophication of a fresh water basin´s data. In “Models and Methods for Supporting Decision Making in Human Health and Environment Protection”, (Bouza et. al Editors). NOVA SCIENCE PUBLISHERS, N.YORK.
2. Brooks, R. (2010):Advanced Derivatives and Strategies. Introduction to Derivatives and Risk Management (8th ed.). Mason, Ohio: Cengage Learning, 483–515.
3. Chen, C.Q.; Tang, S.L.; Xing, Q.G.; Yang, J.K.; Zhan, H.G.; Shi, H.Y. (2007): A Derivative Spectrum Algorithm for Determination of Chlorophyll-a Concentration in the Pearl River Estuary. In Proceedings of Geoscience and Remote Sensing Symposium, Barcelona, Spain.
4. Cheng, C., Y. Wei, X. Sun and Y. Zhou (2013): Estimation of Chlorophyll-a Concentration in Turbid Lake Using Spectral Smoothing and Derivative Analysis. Int. J. Environ. Res. Public Health , 10, 2979-2994 .
5. Le, C.F.; Li, Y.M.; Zha, Y.; Sun, D.Y.; Huang, C.H.; Lu, H. (2009): A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: The case of Taihu Lake, China. Remote Sens. Environ.,113, 1175–1182.
6. Zimba, P.V.;Gitelson, A.A. (2006): Remote estimation of chlorophyll concentration in hypereutrophic aquatic systems: Model tuning and accuracy optimization. Aquaculture, 256, 272–286.
7. Zhang, Y.L.; Qin, B.Q. (2001): Study prospect and evolution of eutrophication in lake Taihu. Shanghai Environ. Sci., 20, 263–265.

Agro-SCADA: An SCADA system to support Sensor Monitoring in Agriculture

Jhon Padilla[1], Jorge Caicedo[2]

[1] Universidad Pontificia Bolivariana
Address
jhon.padilla@upb.edu.co

[2] Universidad Pontificia Bolivariana
Address
calidra@hotmail.com

Abstract

This paper describes an SCADA system (Supervisory Control And Data Acquisition), which was developed to support sensor monitoring in Agriculture applications. This system is based on electronic devices that use MODBUS and Zigbee protocols. Also, our system has a Server that contains the Pentaho suite, which allows to support all the steps required for Decision support. This work describes the technology platfform that we have built and some proofs that we have performed until now. The system that we build is an important platfform to develop other projects for DSS in Agriculture.

1 Introduction

In recent years, new technologies to build Sensor Networks have arised. One of them is the Zigbee technology, that is based on IEEE 802.15.4 protocol. Zigbee plays an important role today in the world of Wireless Sensor Networks (WSNs). Wireless Sensor Networks are composed by a great number of nodes that have wireless data communication modules and sensor modules. By means of this combination, variables as temperature, humidity, CO2 percentage, etc., can be measured and transmitted over the air until Data processing points, which could be Servers that show graphics, statistics, alarms, etc. On the other hand, other technologies that until now have been used in Industrial Data Networks, such as MODBUS networks, now can be used in Agriculture. This technology can be mixed with WSN based on Zigbee to bring support to Agriculture applications that need to obtain measures of variables which are important for several Crops, and thus, they bring information to DSS systems in Agriculture production.

The use of the tools named before for agriculture is in an early stage in the world and their use is not common yet among farmers. With the aim of reduce the digital divide that exists in Latin American agriculture fields, we have developed an infrastructure composed by Controllers and nodes based on communications technologies such as Zigbee, MODBUS and GPRS (the data service for Cellular mobile Networks). Employing such a sensor network, we can sense several variables in crops and then transmit their measures by means of wireless communication methods until a central point that has interconnection with internet, and thus, it is connected to a Data Processing server that is based on data processing technologies that allows to perform data mining and Big Data processing.

In our system [2], we used as software plattform a free version of Pentaho Suite [3], developed by Hitachi Corporation. In this early stage of the project, we can take the data and make graphics of variables versus time. Then, we use an analogy of the Supervisory, Control and Data Acquisition systems (SCADA) used in Industrial Data Networks to give a name for our system (we named Agro-SCADA), because our system is being used in Agriculture environments. In a second part of our project, we will use data mining to discover relationships and to make proyections for typical Colombian crops such as Tomatoes.

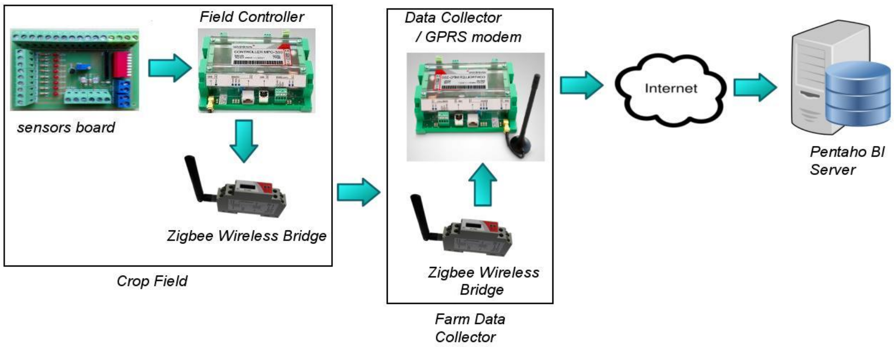


Figure 1: Agro-SCADA system infrastructure

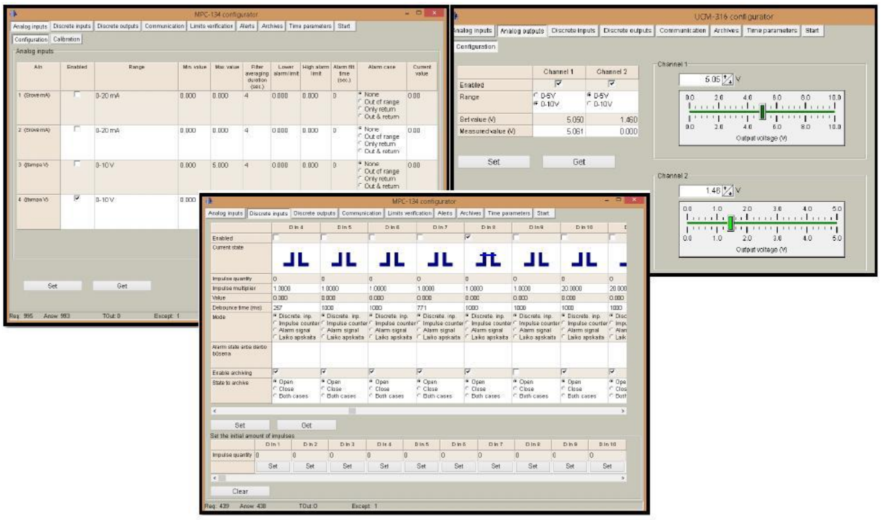


Figure 2: sensors configuration interface for MODBUS Controller

2 Developed infrastructure Description

Agro-SCADA system (Figure 1) is an infrastructure that take measures from crops and transmit those measures to a Data Server that shows graphics for several variables versus time. To do this, it is composed by MODBUS Controllers developed by advanticsys [1], which are allocated in crop fields and they take measures from sensors. At the same time, controllers are connected to Zigbee Wireless Bridges that transport data over the air until a central point, which is a data collector that send that information over a mobile data link (with GPRS technology) to a Pentaho BI Server allocated in any site in Internet.

3 Results

We have developed several tests and we have obtained good results. First, it is possible to configure all the sensors connected to de MODBUS controller in an easy way by means of a graphical interface as can be seen on Figure 2.

Also, we use wireless sensors that measure CO2, relative humidity and Temperature and send those data by means of Zigbee technology until the Farm Data Collector directly or by means of the Wireless Zigbee Bridge.

Besides, we built several topologies for our sensor networks. This is important because we can adapt our system to different situations in several crops. One key feature of our system is that it uses GPRS

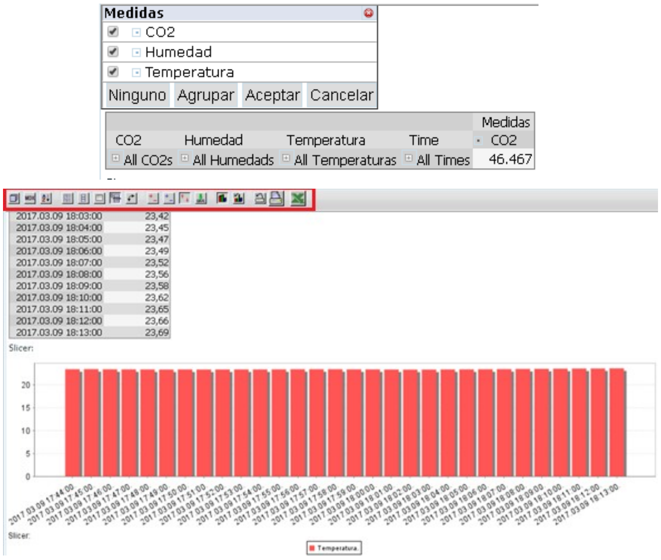


Figure 3: Pentaho BI Server graphical interface

technology to connect to Internet. With this issue, it is possible to connect our system to internet from any place that have cellular phone communications. Thus, it is no necessary to have fixed internet service, which could be a problem in certain zones in the country where cellular communications with GPRS technology are common.

Data are registered in Data Collector Node in hexa code. Such Data are transmitted to the server in a text file form, and then, that file can be showed in Pentaho BI Server as can be observed in Figure 3, where a graph for selected variable is displayed.

References

[1] advanticsys. http://www.advanticsys.com. on line, 2016.

[2] Jorge Caicedo. MANUAL DE CONFIGURACIÓN E IMPLEMENTACION PARA LOS CONTROLADORES UCM-316 Y MPC-134 EN LA APLICACIÓN DE AGRICULTURA. Technical report, Universidad Pontificia Bolivariana, 2017.

[3] Hitachi Corp. http://www.pentaho.com/product/business-visualization-analytics. on line, 2016.

A tree canopy counting method for precision forestry

Juan Pablo Garella[1]*, Matías Tailanián[1], Gabriel Lema[2], Germán Fernandez Flores[1], Javier Regusci[1], Mónica Almansa[1], Pedro Mastrángelo[1], Pablo Musé[3]

[1] CSI Ingenieros, Montevideo, Uruguay
jpgarella@csi-ing.com*

[2] Ecole Normale Supérieure Paris-Saclay, France

[3] Facultad de Ingeniería, Universidad de la República, Uruguay

Abstract

In this paper we present a model-based tree canopy counting method developed for precision forestry applications. The proposed approach is at the core of an in-production service that provides accurate estimates of forestry assets, through the use of unmanned aerial vehicles and a novel combination of image processing techniques. About 10 million trees have been surveyed nationwide, responding to the demand of several major actors in the Uruguayan forestry industry. This massive amount of data opens a new door to investigate the application of machine learning and big data techniques that will certainly improve the overall accuracy and efficiency of the service.

1 Introduction

Exports of forest products (wood, cellulose and paper) currently represent 17% of the total exports of goods from Uruguay [3]. In this context, forest inventory turns out to be of major importance. Over the years several studies have been carried out with the objective of performing individual tree crown analysis and detection over forested areas [1, 2]. To this end different remote sensing techniques have been used, such as satellite imagery [4], airborne LIDAR [7] and aerial photogrammetry [6]. Additionally, drone usage is growing rapidly in precision agriculture and forestry applications [5].

In early 2016 we started to deliver a tree canopy counting service to several major actors in the Uruguayan forestry industry by means of unmanned aerial vehicles (UAVs) technology. Through the use of high resolution aerial imagery acquired by UAVs and the application of specifically designed image processing algorithms, we performed detection, counting and subsequent analysis of the distribution of live and dead, large and small, healthy and diseased trees in forested regions. The remainder of this paper is devoted to present the steps involved in the proposed approach, and to discuss its results and future research directions.

2 Tree Canopy Counting Process

The tree canopy counting process involves the following stages: (1) Image acquisition; (2) Preprocessing. Maps generation: Orthomosaic, Reflectance, NDVI, 3D Models; (3) Automatic tree canopy detection; (4) Quality control and manual correction; (5) Output generation: trees geo-localization and value added products (e.g. tree density maps).

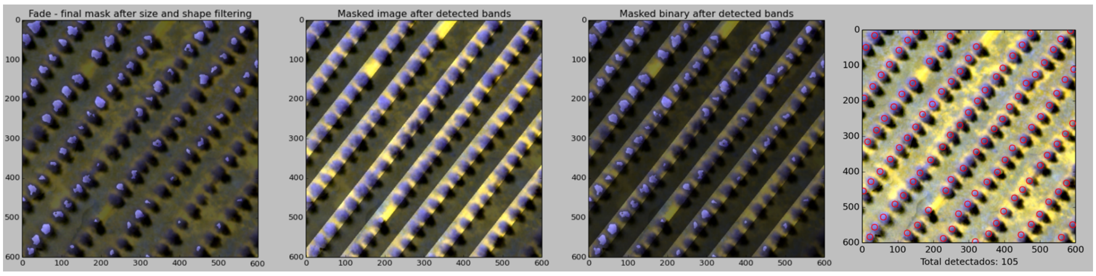
(a) (b) (c)

Figure 1: Pictures of 3 to 6 months old Eucalyptus forested regions

The first stage is performed by using a NIR (Near Infra Red) and an RGB camera mounted on a precision agriculture UAV, the Ebee Ag drone produced by SenseFly (Online, `https://www.sensefly.com/home.html`). Among other advantages, this setup allows us to capture high-resolution aerial photos with a ground resolution of 3 to 4 cm/pixel. The main feature of the service consists in sensing 3 to 6 months old Eucalyptus trees (in between 30 and 50 cm diameter). Accurate detections can only be reached with high-resolution, good quality images. Otherwise, even for human observers the visual recognition of each tree in the forest could be a difficult task. Futhermore, depending on the season and soil conditions, it is quite common to encounter regions where the trees show less contrast with the ground or even significant differences in size. Figure 1 illustrates this point. Nonetheless, it is worth mentioning that in the course of this service older trees with at least twice the size (one meter in diameter and one and a half meters high) were processed as well.

In the second stage we use the acquired NIR and RGB images and the Pix4D software (Online, `https://pix4d.com/`) to create geo-referenced maps of the field.

Third stage is the core of the process; the automatic tree canopy counting algorithm is applied over the output of the previous stage. This is done by splitting the map (i.e.: Reflectance Map) to be used into several overlapped images. The algorithm is applied to each one of these images. To this end, the algorithm parameters have to be calibrated. This is done using a representative set of these images. The parameters include the radius of the trees to be identified, the distance between rows of trees, the distance between trees in the same row, segmentation thresholds, among others. The system gives the option of performing this calibration process semi-automatically by learning the parameters on the aforementioned subset of images. Part of the algorithm is illustrated in Figure 2, with 3 to 6 months old Eucalyptus. In the leftmost image the trees are segmented based on image intensity, size and shape factor using restrictive thresholds to avoid false positives. These trees are then used to detect the plantation layout and to remove regrowths from previous plantation cycles. The remaining trees are the ones considered as true positives. The rightmost image shows the final result of the process. Notice that among all trees that can be identified by visual inspection, only one of them was not detected by the algorithm. The output of this stage is a file with geo-referenced points centered on each detected tree over all the region of interest.



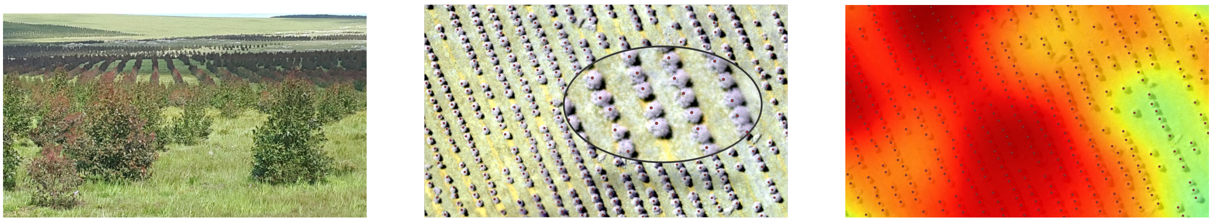
(a)

Figure 2: Illustrative example of the tree canopy counting process.

In the fourth stage automatic detections are subject to visual inspection and manual correction if needed, in order to ensure that the output meets the precision level required by the client.

In the final stage we generate useful information to assist forestry managers in the decision making process. For instance, a density map based on the tree's spatial distribution may allow the forestry workers to improve their efforts exploring the field as well as other related tasks. Figure 3 illustrates a typical outcome of the entire process over a 1 year old Eucalyptus forested region. Figure 3(a) shows a picture of the field; Figure 3(b) shows the orthomosaic map (output of the second stage) over a sample area labeled with the detected trees (overlapped dots); Figure 3(c) shows the corresponding density map. Red areas represent higher densities of trees while green areas correspond to lower densities.

The advantage of this approach over the classic one used in forest inventory is that no extrapolation of manual measurements in the field is needed in order to estimate the overall population of trees. More precisely, the approach that is still widely spread consists in performing a manual survey over a randomly chosen set of regions, computing the average density and then consider this density as the density of the



(a) (b) (c)

Figure 3: Illustration of a typical outcome of the automatic tree canopy detection over a 1 year old Eucalyptus forested region.

whole plantation. This methodology is extremely time consuming and far less accurate.

3 Conclusions and Future work

In this paper we presented a description of our tree canopy counting system. In the course of delivering this service to the forestry industry we surveyed about 10 million trees. This massive and ever-growing amount of data allows us to generate a huge database of geo-localized trees in conjunction with the corresponding index maps generated by the Pix4D software from the UAVs acquired images. To the best of our knowledge, at least regionally, this is the first labeled database of its kind. Having this database at our disposal opens us the possibility to explore data-driven approaches. In the near future we will use this database to train and test machine learning algorithms such as deep neural networks to perform individual tree crown detection tasks in forestry sensing applications.

4 Acknowledgments

This work was partly supported by the Uruguayan National Agency for Research and Innovation (ANII). G. Lema and P. Musé were with CSI during the development of this work.

References

[1] M. Erikson. *Segmentation and classification of individual tree crowns*, volume 320. 2004.

[2] F. et al. Gougeon. Individual tree crown image analysis—a step towards precision forestry. In *Proceedings of the First International Precision Forestry Cooperative Symposium, Seattle*, pages 43–49, 2001.

[3] Uruguay XXI Investment and Export Promotion Agency. Investment opportunities in the forestry sector. `http://www.uruguayxxi.gub.uy/informacion/wp-content/uploads/sites/9/2017/03/Forestry-Industry_Uruguay-XXI_2017.pdf`, February 2017. [Online; accessed 25-May-2017].

[4] Mamoru Kubo and Ken-ichiro Muramoto. Matching of high resolution satellite image and tree crown map. In *Proceedings of ISPRS*, volume 112, pages 1401–1403, 2008.

[5] L. Tang and G. Shao. Drone remote sensing for forestry research and practices. *Journal of Forestry Research*, 26(4):791–797, 2015.

[6] L. Wang. A multi-scale approach for delineating individual tree crowns with very high resolution imagery. *Photogrammetric Engineering & Remote Sensing*, 76(4):371–378, 2010.

[7] Z. Zhen, L. Quackenbush, and L. Zhang. Trends in automatic individual tree crown detection and delineation—evolution of lidar data. *Remote Sensing*, 8(4):333, 2016.

SOC IoT data collection platform: *Application to oceanic temperature sensing*

Ariel Sabiguero Yawelak[1], Angel Segura[1]

Facultad de Ingeniería - Centro Universitario de la Región Este
Universidad de la República
asabigue@fing.edu.uy asegura@cure.edu.uy

Abstract

Coastal ecosystems are highly dynamic and present variability at multiple scales. This variability would be further amplified in a global change scenario due for example to increased runoff, extreme wind events among others. The Uruguayan coastal zone present gaps in their monitoring, precluding the generation of accurate predictive models, analysis of vulnerability, understanding of coastal dynamics and biodiversity changes. Meteorological buoys are the standard approach to this data collection scenario. Despite its relevance for biodiversity management, risk analysis and global understanding of coastal circulation patterns, Uruguayan coastal zone is not adequately monitored. This is mostly because technological developments have been made in Northern hemisphere and the associated cost precludes its instalation and maintenance. The arrival of small, efficient and cheap computer systems, spinoff of the smartphone explosion, is suggesting a new approach to solving several classes of problems with a full computer running a complete operating system, instead of highly specialized embedded systems. This work presents a proposal for the replacement of the classical embedded system with a small system: a Raspberry Pi. The proposal aims at replacing nearly dumb sensors with intelligent ones in a sensor network, introducing concepts of the Internet of Things. Initial results are introduced and subsequent steps for the first buoy implementation are presented too.
Keywords: global warming, coastal sensors network, IoT, Raspberry Pi, SOC

1 Introduction

Deciphering oceanographic patterns in coastal ecosystems is critical to understand and predict natural and anthropogenic induced variability. The increase in average sea surface temperature has profound effects in coastal ecosystems. However, the predicted increase in variability could foster huge changes. Coastal ecosystems provide natural services and support most human activities. The variability of these highly dynamic ecosystems require high frequency monitoring. In that sense, understanding and monitoring the different scale variability of fundamental state variables (i.e. temperature, salinity) of marine-coastal systems is crucial. Uruguayan coast is an area of high oceanographic complexity and great space-temporal variability due to the dynamic interaction between coastal streams of the currents of Brasil and Malvinas and the influence of freshwater from the second biggest estuary of South America: the Rio de la Plata. This variability is unique and presents patterns at different time-scales. Characterization of these by means of autonomous sensors has proven an effective tool that allows real-time gathering of obtained data. Data can be used to feed physical models, generate forecasts that help to understand the behavior of these complex systems [2, 4]. Its knowledge would lead to a more informed management of the coast area. Temperature data collection with meteorological buoys is already taking place, but for several reasons, a new approach is required. The buoys are expensive devices, and costs are always something that needs to be addressed, as savings can be turned into more data series and more research. Buoys are not produced in our country, and time issues also affect research quality. The environment where buoys operate is a difficult one, and they also present failures. Time involved in fixing a failure may involve up to 6 months worth of data not collected. Fixing a buoy requires fetching it from the sea, moving it to laboratory detecting the failure, export the failed component to factory, get the replacement, importing it back and re-deploying it in the sea. This work presents a proposal for a small computer, IoT, commodity components based that addresses a data gathering platform.

2 Problem description

The initial objective is the placement of data gathering units on five existing buoys over a coastal extension of 150Km. The buoys are located from 1 to 3 Km away from the coast. Measurements of the water temperature must be taken 1m below the surface every hour. The information should be accessible immediately. Data gathering must be priorized over transmission.

2.1 Proposal

It is important to note that there is no relevant challenge regarding the data gathering and its transmission inside the laboratory. Turning a Raspberry into a weather station has already been addressed. Existing proposals mostly keep the Raspberries indoors or in a controlled environment, connected to the mains. The challenge is to turn this indoor device into a solution that can be deployed in a standalone, autonomous unit, able to operate reliably in a hostile environment.

The main identified challenges are: energy, data transmission and physical preparation for the environment.

2.2 Aspects related to energy

State of the art in electronics does not allow to reach equivalent levels of energy consumption using a SOC (system on chip) than with a microcontroller. The development of the solution based on a microcontroller, the classical approach, is a solid one, but costs are pushed to the development phase. Use of a general purpose SOC, with a state of the art GNU/Linux operating system, makes available every option to the developer. The power of the platform is unprecedented. Tasks like data preparation, validation and correction can be pushed to the edge of the sensor network, as the sensor is capable of 2451 MIPS.

There are at least three main sources of energy: solar, wind and sea currents. The buoy is fixed to the seabed, thus, relative movement of air and sea can be used to produce electricity. As both sources require moving parts and hazards, they will not be considered for the initial prototype, that will only use solar energy. Considering monthly irradiation data published in [5] and the data available from the Laboratorio de Energía Solar - LES the buoy will be exposed to at least $8MJ/m^2$ a day on average during winter. Considering state of the art (20% efficient or more) solar panels with adequate surface and conservative power saving measures, the system should get enough energy to operate continuously, maybe, transmitting once or twice at night or in cloudy days. Efficiency of the whole solution has to be optimized at every level.

2.3 Aspects related to communications

The location of buoys, from 1 to 3 Km away from the coast, gives a high level of confidence that every data gathering point can be connected to the cellular network. According to coverage maps from Uruguay's national telephone operator there is complete coverage of the coast and the existence of several seaside resorts with 4G coverage promises adequate connection availability. We were able to verify existence of cellular link with our own mobile phones on some of the locations. Details on network quality and how to address them will appear after the initial prototype is deployed.

A cellular network device uses little power when the signal is strong, but as signal fades, it increases its transmission power, hence, the power consumption. The transmission frequency can be adjusted to energy production and signal strength in order to optimize running time of the batteries and still meeting transmission requirements. Between transmission events, the device should not only be disconnected from the network, but, power should be removed from the modem, as it employs energy to be connected to the cellular network even if no data connection is established.

We would like to have the buoy on-line all the time possible. If energy production exceeds consumption and batteries are fully charged, there is no need for energy saving, allowing real time access to data gathered. It is relevant to note that this is also the moment where security becomes an issue. If search engine bots indexes the buoy, it receives port scans or dictionary attacks or people consider it interesting reading its temperature, they draw energy. General IoT security analysis have to be done, as DoS in this scenario might mean exhaustion of battery power and the device not able to perform its task.

2.4 IoT architecture

Internet of Things proposes ways to introduce processing power to the edges of sensor networks. Different researchers [1, 3] propose different alternatives for task organization and complexity distribution. Despite the different proposals, most proposals suggest that Data and Presentation Layers should be implemented at the buoy level. Transversal layers, like Security, must be somehow implemented too. Architectural discussion involves the level of implementation of the Business Layer, dominant factor concerning energy usage for data transmission. Business requirements must be modeled at the buoy level in order to compete with energy saving rules. Presentation layer must be discussed, but it seems that it must be implemented at central level. Buoys might not be connected at all times, moreover, they are unreliable for data storage, as many accidents put in risk their ability to survive events like collision with vessels or big mammals. Data must be collected centrally, and offered on a 24x7 basis for research purposes. Streaming services for analytics systems must be provided too. All these requirements, that might be provided from the buoy if we only consider processing power, must be offloaded considering energy consumption.

3 Expected results

Different groups at CURE (Centro Universitario de la Región Este) cooperate in order to produce the hardware, code the software and use the data for scientific modeling. Short term objectives involve having a working prototype, floating near research facilities in order to experiment with different power optimization algorithms and learning about deploying SOC in these environments. Mid term objectives involve an optimized solution, running in all five selected buoys, processing and sending data. These data must be gathered and compared to satellite models and imagery available, and introduced into biological models of the coast region. The availability of this powerful platforms open the opportunity of developing additional sensing capabilities. They could be offered to communications, computer science and electronic careers in our university as platforms for algorithm optimization, sensor development and the development of new models for oceanographic research.

References

[1] Leonardo Albernaz Amaral, Everton de Matos, Ramão Tiago Tiburski, Fabiano Hessel, Willian Tessaro Lunardi, and Sabrina Marczak. *Middleware Technology for IoT Systems: Challenges and Perspectives Toward 5G*, pages 333–367. Springer International Publishing, Cham, 2016.

[2] A.Patynen, J.A.Elliott, P. Kiuru, J. Sarvala, AM Ventela, and R.I. Jones. Modelling the impact of higher temperature on the phytoplankton of a boreal lake. *Boreal Environment Research*, 19(1):66 – 78, 2014.

[3] Martin Bauer, Mathieu Boussard, Nicola Bui, Jourik De Loof, Carsten Magerkurth, Stefan Meissner, Andreas Nettsträter, Julinda Stefa, Matthias Thoma, and Joachim W. Walewski. *IoT Reference Architecture*, pages 163–211. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[4] Hans W. Paerl and Valerie J. Paul. Climate change: Links to global expansion of harmful cyanobacteria. *Water Research*, 46(5):1349 – 1363, 2012. Cyanobacteria: Impacts of climate change on occurrence, toxicity and water quality management.

[5] R. Alonso Suárez, G. Abal, P. Musé, and R. Siri. Satellite-derived solar irradiation map for uruguay. *Energy Procedia*, 57:1237 – 1246, 2014. 2013 ISES Solar World Congress.

Design of a low power wireless sensor network platform for monitoring in citrus production

L. Steinfeld, J. Schandy, F. Favaro, A. Alcarraz, J. P. Oliver, F. Silveira

Instituto de Ingeniería Eléctrica, Universidad de la República
Julio Herrera y Reissig 565 - Instituto de Ingeniería Eléctrica
{leo, jschandy, ffavaro, alcarraz, jpo, silveira}@fing.edu.uy

Abstract

Wireless sensors networks (WSN) enable the acquisition of valuable data directly from the production field that, in turns, open unprecedented opportunities for data analysis and decision making systems. Monitoring microclimate variations, typical in citrus crops, could allow precise irrigation scheduling and harvest planning. However, a limited lifetime of battery-powered sensor nodes may impose barriers to its widespread adoption. This work presents the overall architecture and main characteristics of a WSN solution based on open standards and free software. The main techniques applied to reduce power consumption are described, obtaining an expected node battery lifetime of more than two years.

1 Introduction

Informed management of agricultural production requires distributed data of relevant conditions across the producing fields. Advances in embedded systems and communications technologies make possible to monitor distributed measurement points with the following characteristics. Wireless nodes are low cost enough so that a dense measurement array can be deployed. Two additional characteristics are key for providing an easy to use service to agricultural technicians and producers. The nodes need to have very low power consumption so that they operate during long periods (several months to more than one year) without changing batteries or by harvesting energy from the environment. The hardware, software and the communication protocols need to be reliable and robust so that seamless network setup and unattended operation are possible. Several wireless sensor networks (WSNs) have been applied to precision agriculture for irrigation systems [1, 2], and frost detection [3].

This work presents the overall architecture and main characteristics of a complete solution being tested in an actual production context. The proposed solution is based on open standards and free software and the main techniques applied for autonomous, low power operation are summarized. The solution is being tested in a citrus farm for the following goals. First, irrigation monitoring aiming at efficient use of water and energy. Second, monitoring of microclimate conditions, mainly for frost detection. In the case of citrus, production fields with significant topographic variations are common. This leads to irrigation and microclimate variation along the production field. Both have important consequences on how the production is handled, either irrigation management decisions or harvest decisions (related to frost impact) or even extent of pesticide use. The presented system provides the producer with timely and detailed information for decision making based on data available online. A pilot network is being deployed in a citrus orchard in Margat, Canelones in the south of Uruguay.

2 Design

Fig. 1 depicts the overall system architecture. Sensor nodes form a wireless ad-hoc network based on IEEE 802.15.4 operating in the unlicensed 2.4 GHz band. Sensor nodes measure and report environment data to the network root node of the gateway. The gateway has both IEEE 802.15.4 and 3G connectivity, thus send the sensor information to a remote server in Internet via cellular network. Remote users

This work was performed under INIA FPTA grant 313, Project: Gervarsio.

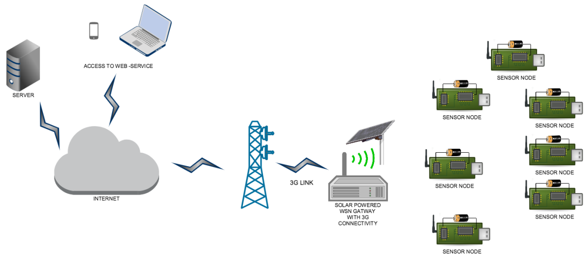


Figure 1: System architecture

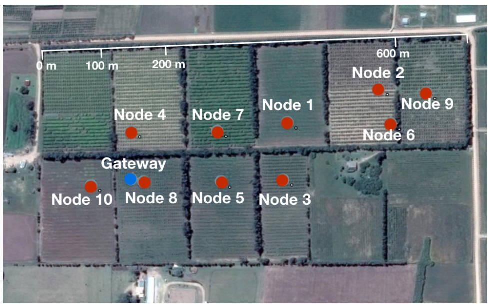


Figure 2: Image of citrus crop and sensors locations.

can access the information from their cellphones or personal computers by accessing a web service in Internet. Fig. 2 shows the actual location of sensors nodes and the gateway. The network currently has ten sensor nodes, deployed one per frame, each about one ha, thus the distance between sensors is roughly 100 m.

2.1 Hardware

The sensor node core is a CC2538 System-on-Chip (SoC) manufactured by Texas Instruments, which integrates an ARM Cortex-M3-based microcontroller with an IEEE 802.15.4 radio. The EMB-Z2538PA module by Embit was used for the PCB custom made design, which includes a CC2538 and a CC2592 PA/LNA front end delivering a RF output power up to $+20$ dBm. The node is powered by two AA lithium-ion or standard Alkaline batteries in series, supplying a nominal voltage of 3.0 V. We adopted the TPS62740 step-down DC-DC converter to reduce the node power consumption, which achieves efficiencies greater than 90% for a current drawn as low as $10\mu A$. This guarantees savings even during very low power drain, typical of the microcontroller low-power modes.

The node is equipped with the following sensor set: soil humidity (Decagon EC-05), air temperature and humidity (SHT-21 from Sensirion) and soil temperature (TMP275 from Texas instruments). The node and antenna are packaged in a IP65 case that is easily mounted on a pole made with a standard galvanized water pipe to connect with sensors in a meteorological shield and sensors on the ground. The gateway includes a single-board computer (Raspberry PI (RPI) model B+ v1.2), the sensor network root node and a 3G cellular modem. The gateway power supply system is comprised by a 50W solar-panel, a 12V 24Ah VRLA battery, and a solar charge controller. The photovoltaic system is designed to endure a couple of null solar generation days (cloudy sky in winter) for the worst case power consumption (low signal strength in 3G modem), providing a high autonomy even in bad climate conditions.

2.2 Embedded software and communication stack

Contiki is an open source, event-driven operating system oriented to WSN and IoT applications using constrained hardware. Contiki manages the hardware resources and includes different libraries such as network stacks. The Contiki distribution includes the network protocols' stack most widely used in WSN, described next. The physical and MAC layers are based on the standard IEEE 802.15.4. 6LowPAN is an adaptation layer protocol by the IETF (RFC 4944 and 6282) that allows the transport of IPv6 packets over 802.15.4 links. It is in charge of the compression of IPv6 and the upper layer headers and of the fragmentation and reassembly of IPv6 packets. RPL is the adopted routing protocol based on a tree-oriented strategy (RFC 6550), in which nodes join the network dynamically forming a mesh, and traffic flows to a root node.

The Constrained Application Protocol (CoAP), at the application layer, is a RESTful protocol for use with constrained hardware such as WSN nodes, since uses UDP underneath. The REST model works with server nodes that make certain resources available under a URL, allowing to have a client / server architecture based on a standard protocol. Client nodes access resources using methods such as GET, PUT, POST, etc. In this work we use the OBSERVE mechanism, which allows client nodes to retrieve a resource value from a server (GET) and keep it updated over a period of time.

The overall architecture, which is based on widely used standards and open protocols, allow to take profit of several freely available tools (e.g. simulator, protocol analyzer, framework).

The gateway runs an embedded Linux operating system (Raspbian Jessie distribution), that allows to execute the scripts or daemon that rely the wireless sensor network data to a server through Internet. The server side of the application and user software is treated on a companion paper.

2.3 Low power design techniques

A very low power consumption is achieved thanks to several design decisions. In the communication stack, the use of 802.15.4 with duty cycling through the ContikiMAC protocol allows the system to be over 97% of the time in low power mode. Also the use of appropriate protocols at higher layers (e.g. CoAP) allows to keep low power consumption. Regarding the node power management, Contiki takes full advantage of the microcontroller low power modes, powering down the microprocessor when there is neither processing needed nor events scheduled in the event queue.

When the sensors are idle, they are turned off through low leakage switches to lower its power drain. Additionally, the supply voltage of the node is lowered to further reduce the overall power consumption. The microcontroller optimal supply voltage to minimize the power consumption is 2.1V, but some sensors require a minimal power supply of 2.5V. The selected DC-DC converter has selectable output voltage. The microcontroller dynamically control, using an output pin, the supply voltage. While sensors are active and measuring the supply voltage is selected to 2.5V, and remaining time to 2.1V. As a result, the microcontroller active current is reduced from 1.3mA @ 2.5V to 1mA at 2.1V. With this technique, we also avoided power hungry level shifter stages to interact with the sensors. Finally, a careful setting of the I/O state of microcontroller pins in order to minimize consumption during sleep mode due to open inputs and pull ups / pulls downs. This performance gives an expected node battery longevity of more than two years in a leaf node that transmits sensor data every 15 minutes powered by batteries with 2.8Ah useful charge.

On the gateway side, the 3G modem is connected to the RPI via USB through a specially designed circuit that allows switching the power supply of the modem. Turning off the modem when it is not transmitting (75% of the time by design) results in a 16.5% reduction in the overall power consumption. This also acts as prevention in case of a potential hung up of the modem.

3 Conclusion and future work

The overall architecture of a WSN infrastructure for precision agriculture was presented. A reliable operation and reduced power consumption is mandatory to run unattended for years. This work summarized the main techniques applied to extend sensor node's battery longevity up to an expected lifetime of two years. A pilot was deployed in a real citrus orchard for testing purposes and subsequent validation. This kind of platform constitutes the basic infrastructure at the very early stage of a future decision support system.

References

[1] O. Adeyemi, I. Grove, S. Peets, and T. Norton. Advanced monitoring and management systems for improving sustainability in precision irrigation. *Sustainability*, 9(3):353, 2017.

[2] J. A. López, F. Soto, J. Suard'ıaz, P. Sánchez, A. Iborra, and J. A. Vera. Wireless Sensor Networks for Precision Horticulture in Southern Spain. *Comput. Electron. Agric.*, 68(1):25–35, Aug. 2009.

[3] F. J. Pierce and T. V. Elliott. Regional and On-farm Wireless Sensor Networks for Agricultural Systems in Eastern Washington. *Comput. Electron. Agric.*, 61(1):32–43, Apr. 2008.

Development of a wireless sensor network system for the monitoring of insect pests in fruit crops.

Dpto. de Electrónica y Dpto. de Procesamiento de Señales

Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República
Julio Herrera y Reissig 565 . Corresponding authors [silveira, lbarboni, agomez]@fing.edu.uy

Abstract

Wireless sensor networks are becoming an essential technology in precision-agriculture and environmental monitoring. Their low cost and autonomous functioning enables distributed and scalable deployments in extended domains such as farms. Valuable information from the fields can then be captured and transmitted automatically allowing intelligent decision making. In an ongoing collaboration between the University and fruit producers, a wireless sensor network system is being developed for the monitoring of insect pests.

1 Introduction

Wireless Sensor Networks (WSN) are comprised of small programmable devices named nodes. Each node is composed of a micro-controller, sensors and a radio to communicate wirelessly with neighboring nodes. Their increasingly widespread usage is due to the following reasons: *i)* the node low-cost enables to build distributed deployments easily scalable with large spatial density of nodes per unit area with reduced cost of installation and maintenance, *ii)* these deployments enable to build maps describing scalar fields varying in time and space (e.g. temperature, humidity), *iii)* the nodes operate with low current consumption, so that they can achieve several months of battery lifetime.
Most WSN applications for agriculture have been traditionally restricted to scalar measuring nodes, see for example [1, 2], but recently image capturing capable nodes are being integrated with the challenge of handling more complex data over the networks. Some works in this line are presented in [3, 4]. WSN were not designed to transmit large amounts of data. The transmission of images implies to adapt the use of the existing network protocols or to modify them. In this scenario, well established network protocols and applications already implemented in the real time operating system can be used to collect data in tree or mesh topology networks. In this way, WSNs enable to manage the farm productivity, allowing product quality enhancement with reduced operational cost.

1.1 The problem of insect pests in fruit production

The lepidopterous insect pest (moths) are an important concern in production of fruits such as apples. The moths lay eggs from which larvae are born and they produce lesions to the fruit. The control of the pest population is implemented by means of using plastic traps with a sticky bottom side and pheromone lures (see figure 1). The trap can capture male adult moths attracted by the female pheromone lures. A person, who periodically travels through crops, is in charge of performing the counting of insects caught in the trap and eventually clean trap bottoms of pest crowded traps.

2 Development of the decision support system.

In an ongoing collaboration between the University and fruit producers, a wireless sensor network system is being developed for the monitoring of insect pests. In this system a wireless node is attached to each trap. The node is equipped with a camera that allows capturing images of the trap sticky bottom. The acquired images at each node (average size around 150 kB, 1600x1200 pixels, JPEG format) are forwarded through the network to a central node (sink node). The images are transmitted fractioned into packets and the network protocol ensures that the packets can be correctly assembled at the sink node. Figure 2 presents a simplified diagram of the wireless sensor network system and figure 3 shows the designed image sensor trap.

(a) Left: lesions produced by the larvae. Right: moths.

(b) Typical traps used in apple trees.

(c) Sticky bottom with captured insects.

Figure 1: The problem of insect pests in apple trees and traditional monitoring with traps.

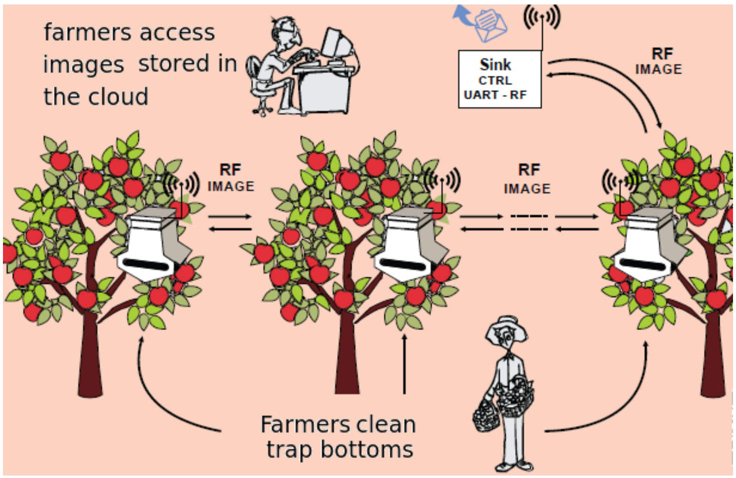


Figure 2: The system for insect monitoring and early warning for plague prevention.

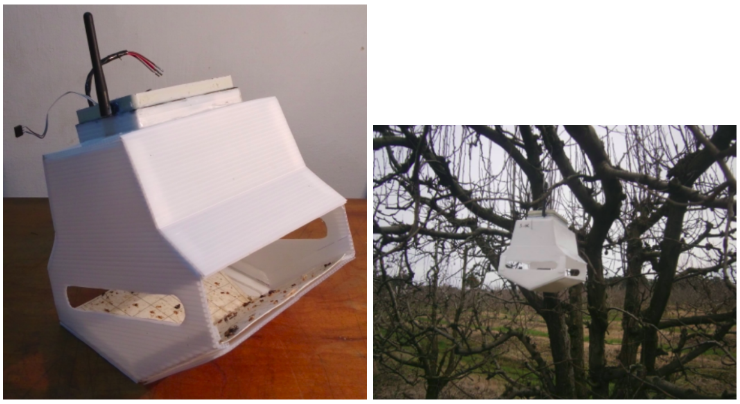


Figure 3: The designed trap with the image sensor node.

The images that arrive to the sink node are transmitted by cellular data connection to a server through a solar-energy powered gateway. Then they are stored and analyzed by an expert. The expert classifies and draws the outline of each object in the image and all the information is stored in a database. With the collected data, several reports of the evolution of the pests can be generated. In the first stage of the project all the images will be analyzed by experts. The manual labeling of the insects in the images by experts allows building a useful dataset that will be used in a next stage of the project to learn how to automatically classify and count the insects. Figure 4 shows the software used by the experts.

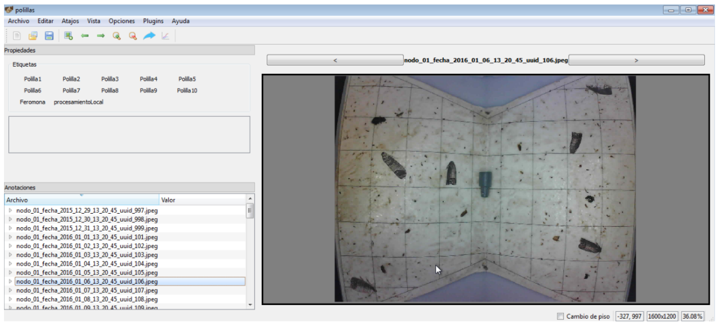


Figure 4: Software used by the experts to classify and outline the insects.

Learned classifiers enable also to move the processing to the nodes as explored in [5]. With the distributed processing, the insects can be detected at the trap node and this information transmitted in a few packets instead of transmitting hundreds of packets with the complete image.

3 Conclusion

A wireless sensor network is being designed to help monitoring pests in fruit plantations. The final system will: i) enable simple pest monitoring of large areas, ii) simplify the maintenance of the traps since the person in charge will only be required for trap cleaning when needed, iii) enable early alerts in case of pest infection allowing localized fumigation with the desired reduced environmental and water pollution.

Acknowledgement

Work performed under INIA FPTA grant 313, Project: Gervasio. To the many valuable contributions made by: F. Silveira, A. Gomez, L. Barboni, B. Marenco, A. Rodríguez, F. Favaro, J.P. Oliver, A. Alcarraz, M. Siniscalchi, A. Vignone, M. Martinez, I. Abadie, F. Arbío, F. Lopez, M. Pereyra, J. Schandy, M. González, N. Wainstein, M. Bertrán, N. Martínez, C. Croce, Cooperativa Jumecal and J. Vila.

References

[1] FJ Pierce and TV Elliott. Regional and on-farm wireless sensor networks for agricultural systems in eastern washington. *Computers and electronics in agriculture*, 61(1):32–43, 2008.

[2] Juan A López, Fulgencio Soto, Pedro Sánchez, Andrés Iborra, Juan Suardiaz, and Juan A Vera. Development of a sensor node for precision horticulture. *Sensors*, 9(5):3240–3255, 2009.

[3] George Nikolakopoulos, Dionisis Kandris, and Anthony Tzes. Adaptive compression of slowly varying images transmitted over wireless sensor networks. *Sensors*, 10(8):7170–7191, 2010.

[4] E. Goldshteina, Y. Cohena, A. Hetzronia, Y. Gazitb, D. Timarb, Y. Grinshpona L. Rosenfelda, A. Hoffmana, and A. Mizracha. Development of an automatic monitoring trap for mediterranean fruit fly (ceratitis capitata) to optimize control applications frequency. *Computers and Electronics in Agriculture*, 139:115–125, 2017.

[5] Mauricio González, Javier Schandy, Nicolás Wainstein, Martín Bertrán, Natalia Martínez, Leonardo Barboni, and Alvaro Gómez. *A Wireless Sensor Network Application with Distributed Processing in the Compressed Domain*, pages 104–115. Springer International Publishing, Cham, 2014.

SIMAGRI: An Agro-climate Decision Support Tool

Eunjin Han[1], Walter E. Baethgen[2], Julieta Souza[3], Mercedes Berterretche Adaime[4], Gonzalo Antúnez [5], Carmen Barreira[6], Flora Mer[7]

[1] Columbia University
116th and Broadway, NY 10027, USA
eunjin@iri.columbia.edu

[2] Columbia University
116th and Broadway, NY 10027, USA
baethgen@iri.columbia.edu

[3] Ministerio de Ganadería Agricultura y Pesca
1476 Constituyente, Montevideo 11200, Uruguay
jsouza@mgap.gub.uy

[4] Ministerio de Ganadería Agricultura y Pesca
1476 Constituyente, Montevideo 11200, Uruguay
mberterretche@mgap.gub.uy

[5] Fundación Julio Ricaldoni
2310 Benito Nardone, Montevideo 11300, Uruguay
mgap.gantunez@ricaldoni.org.uy

[6] Fundación Julio Ricaldoni
1476 Constituyente, Montevideo 11200, Uruguay
mgap.cbarreira@ricaldoni.org.uy

[7] International Research Institute for Climate and Society
km 10 Ruta 48, Rincon del Colorado, Canelones 15900, Uruguay
flora@iri.columbia.edu

Abstract

A decision support tool, SIMAGRI was developed to translate climate information (historical and operational seasonal climate forecasts) to relevant information for supporting strategic and tactical decisions in crop production, e.g., crop choices, adaptive cultural management practices, insurance needs, among others. Two versions of SIMAGRI with different levels of complexities can be used for different purposes (e.g., initial assessment or operational use for generating crop yield outlook etc.). The SIMAGRI tool offers potential for further improvement based on communicating with stakeholders and identifying their demands.

1 Introduction

Managing climate-related risks is one of the key components for enhancing resilience and productivity in agriculture under an increasingly variable climate [1, 2]. Successful agricultural climate risk management can be achieved when climate information is translated into actionable agronomic terms. Therefore, agro-climate tools are required to link climate information to agricultural decision making. The SIMAGRI (Simulador de Agricultura) is an agro-climate tool developed to support agricultural decision making based on historical climate data, seasonal climate forecasts (SCF) and crop simulation models included in the "Decision Support System for Agrotechnology Transfer" (DSSAT). Depending on the climate data that the users select to use, two versions of the agricultural decision support system are available: 1) a simplified DSSAT crop simulation tool for "what-if" analyses based on historical weather observations, and 2) a flexible crop simulation tool for "what-if" analyses using probabilistic SCF. To process flexible "what-if" analyses for informing recommendations, the tool requires large amounts of data and processes, which can be qualified as Big Data Science Process. In this paper, we briefly introduce the two different versions of the decision support tool, SIMAGRI.

2 Methodology

2.1 DSSAT crop simulation

The DSSAT is a modular-based application package of various crop models that can simulate 16 different crops [3]. The DSSAT models simulate daily growth and development of a crop over time, as well as daily changes in the soil water, carbon and nitrogen under specific management practices at a spatially uniform field. Weather data including daily maximum and minimum air temperature (T_{max} and T_{min}), solar radiation and precipitation are fundamental forcing variables to simulate hydrological processes and crop phenology in DSSAT models. The current SIMAGRI is customized for three major crops (maize, soybean and wheat) in Uruguay and runs CERES-Maize, CERES-wheat and CROPGRO-soybean models internally.

2.2 Decision support system based on climatology

Uncertainty arising from climate variability can be considered in crop modeling when long-term historical weather observations are used. Cumulative probability distributions of the simulated crop yields can inform uncertainty due to climate variability. Assuming that each year in the past has equal probability of happening in the future, and that weather statistical properties do not change in the future, the simulated yield distribution can provide useful information for decision-making.
This first version of the SIMAGRI uses long-term historical data from 5 weather stations (Tacuarembo, Salto Grande, Las Brujas, Treinta y Tres and La Estanzuela) to represent different climatic zones across Uruguay. This version is implemented as a web version (http://simagri.snia.gub.uy/webapp/) for an easy-access as well as the original desktop version.

2.3 Decision support system based on probabilistic seasonal climate forecast

When skillful seasonal climate forecasts are available for a coming season, it can be useful to consider the SCF information rather than relying on climatology. Since Uruguay's climate is sensitive to ENSO events and skillful seasonal climate forecasts are available for the country, it may be beneficial to consider the SCF information for agricultural decision making. One of the obstacles to link SCF to a crop simulation model, however, is a format mismatch: SCFs are provided in a probabilistic format (i.e., probability of below-, near- or above-normal category) for the coming 3 ~ 6 months, while the DSSAT crop simulation models require weather inputs at daily time step. To tackle this issue, the second version of the SIMAGRI is equipped with temporal downscaling tools to convert a probabilistic SCF to daily weather sequences and thus enables multiple (e.g., 100 times) DSSAT simulations for a given SCF. Two temporal downscaling methods are currently available within the SIMAGRI: a conditional stochastic weather generator called predictWTD [4] and a non-parametric resampling method called FResampler1 [5].

2.4 Linking Climate data to DSSAT through SIMAGRI user interface

Although DSSAT has been proven as a very useful tool to assist sustainable agricultural management, it can be very challenging for non-experts to apply it for agricultural decision making in operational mode, especially in dealing with uncertainties of future climate. The SIMAGRI allows user-friendly operation of DSSAT by circumventing time-consuming (pre- and post-) data processing and seamlessly integrating it with temporal downscaling methods and SCF.

Convenient Graphical User-Interface (GUI) of the SIMAGRI assists DSSAT simulations for various "what-if" scenarios with different climate forecasts or crop management options. Thus, the SIMAGRI interface has the capability of using the crop models with different cultivars, varying simulation periods from climatology (e.g., last 50 years vs. recent 10 years), various dates/amount/material/methods for planting, fertilizer application, irrigation, soil profiles, etc. The GUI also allows users to query the simulated model outputs and make plots (e.g., box-plots or cumulative probability curves) comparing the results with different management options. Moreover, SIMAGRI integrates crop models with economic data and therefore translate crop model outputs into economic terms. When the user provides information on expected crop prices and input costs (fertilizer, irrigation and general), a simple economic analysis is also possible to explore the gross margins of different scenarios.

The GUI of SIMAGRI was developed using Tkinter module and Pmw megawidgets of the Python script language. The SIMAGRI has flexibility to be updated or to add other functionalities or options in the future. It can be easily customized for different needs (e.g., different crop types), for climate risk management in other regions. In addition, since Python is a free and open-source software, the SIMAGRI can be adopted easily without a license issue.

Figure 1 shows overall flowchart of SIMAGRI version 2 linked with SCF.

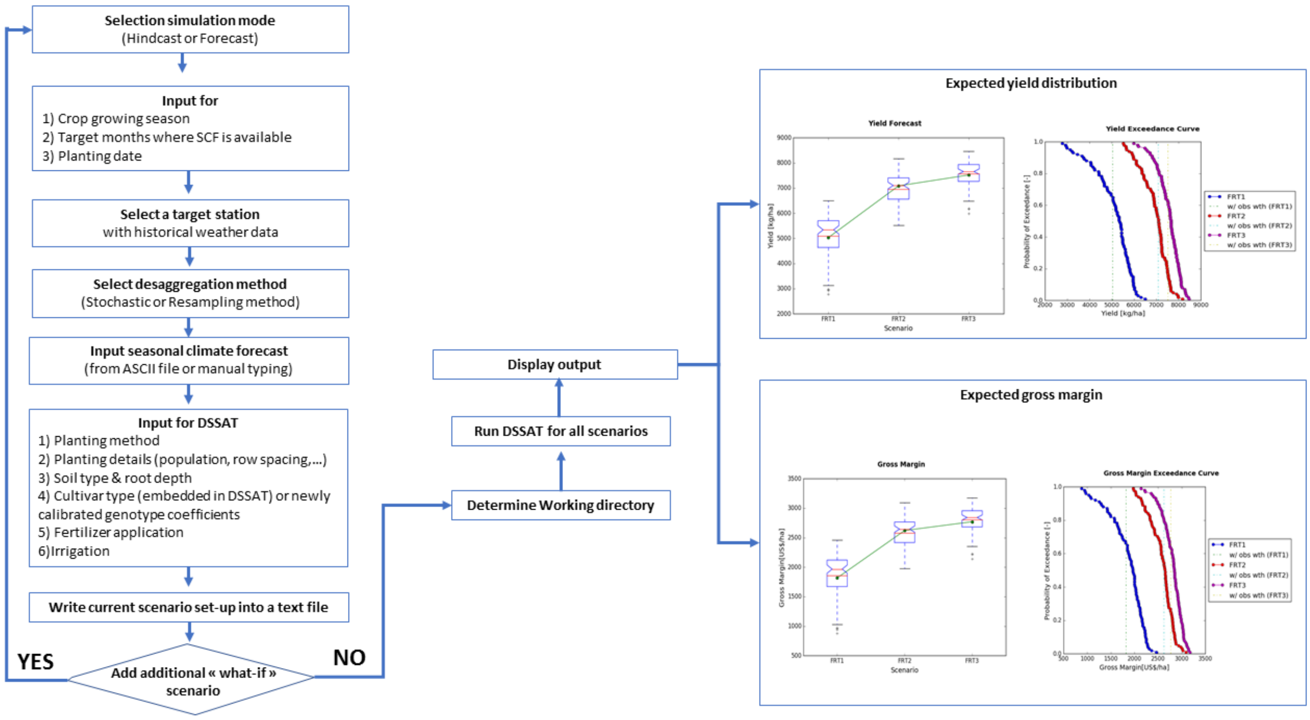


Figure 1: Diagram of SIMAGRI workflow (version 2 linked with SCF)

The SIMAGRI tool was developed by the International Research Institute for Climate and Society (IRI), Columbia University, and is currently implemented as a web tool by the SNIA (Sistema Nacional de Información Agropecuaria), Uruguay.

3 References

[1] Hansen, J. and K. Coffey, *Agro-climate tools for a new climate-smart agriculture.* 2011.

[2] Palombi, L. and R. Sessa, *Climate-smart agriculture: sourcebook.* Climate-smart agriculture: sourcebook., 2013.John Smith. How to make citations. *Journal of Modern Bibliometrics*, 1:1–10, 2010.

[3] Jones, J.W., et al., *The DSSAT cropping system model.* European journal of agronomy, 2003. **18**(3): p. 235-265.

[4] Hansen, J.W. and A.V. Ines, *Stochastic disaggregation of monthly rainfall data for crop simulation studies.* Agricultural and Forest Meteorology, 2005. **131**(3): p. 233-246.

[5] Ines, A.V., *FResampler1: A resampling and downscaling tool for seasonal climate forecasts*. 2013: IRI/Columbia University., NY, USA.

A Decision Support System for Fish Farming using Particle Swarm Optimization

Ángel Cobo[1], Ignacio Llorente[1], Ladislao Luna[1], Manuel Luna[1]

[1] University of Cantabria
Research Group in Economic Management for the Primary Sector Sustainability, Fac. Economics, Avda Los Castros s/n, 39005 Santander (Spain)
angel.cobo@unican.es, ignacio.llorente@unican.es, ladislao.luna@unican.es, manulg1@gmail.com

Abstract

This work presents a DSS for fish farming in cages that determines the production strategies that maximise the present profits of the farming process. It is a model-driven cooperative DSS in which the cultivation process is simulated through a bioeconomic model that considers economic, environmental, biological and technical data. The optimal production strategies are determined using Particle Swarm Optimization (PSO). The DSS helps to deal with two types of decisions: operational decisions in which the DSS obtains the optimal solution under particular conditions; and strategic decisions, in which the DSS allows to obtain the optimal economic result in different scenarios.

1 Introduction

In recent decades, aquaculture has become a relevant industry around the world that is able to sustain the demand for fish. Similar to other animal breeding industries, the management of fish farming is complex due to the broad range of internal and external factors that influence aquaculture, and the complex interactions of technical, biological, environmental and economic aspects during the cultivation process. The rapid development of the aquaculture industry has increased the need for more efficient and productive management systems. These systems have been developed using operational research (OR) methods in order to support managers in decision-making processes. The OR models applied in aquaculture have been based on accumulated experience in fishing and other primary sector activities, such as agriculture or forestry, to increase the efficiency and profitability of fish farming on an industrial scale [2]. The complexity of decision-making for an aquaculture facility suggests the need for computerized analytical tools that can integrate biological, physical, environmental, economic and social components of knowledge required to arrive at a decision [3]. In a broader sense, DSSs address the problem of packaging a large domain of scientific and technical knowledge into a form that is of practical value to a diverse audience, including non-scientists [10].

In aquaculture production is common that managers have limited ecological and biological expertise, which often makes it difficult for them to understand the scientific implications of the production process. Accordingly, DSSs in aquaculture has been mainly directed towards the integration of environmental [6] and biological [4] issues in decision-making processes. Some others have taken other technical aspects into account, such as site selection [9], facilities design [7], managing hatchery production [12], or facilitating aquaculture research and management [4]. However, although aquaculture is an economic activity, this aspect has not been much considered by DSSs. Bolte, Nath, and Ernst [3] developed a DSS for evaluating economic impact in ponds. Also Halide, Stigebrandt, Rehbein, and McKinnon [8] developed a DSS for sustainable cage aquaculture that enables managers to perform an economic appraisal of an aquaculture farm at a given site. The review of previous literature shows the need to develop decision support tools to help achieve economic sustainability of the activity. In this context, the development of a DSS for optimise the economic performance in aquaculture is an innovation in the management of aquaculture facilities that would improve the capability of managers to solve problems and responds quickly and efficiently to changes in the company environment.

2 DSS components

Three fundamental components of a DSS architecture are the database (or knowledge base), the model (i.e., the decision context and user criteria), and the user interface. The knowledge base of the DSS designed includes economic, environmental, biological and technical data, so the system integrates one relational database. This database includes information about locations and environmental conditions (water temperature), biological and economic data of the fish (available fingerlings, minimum commercial size, sale prices, price seasonality…), available feeds and their feeding and growth rates, and technical data of the farm and the cages used in the production process.

The bioeconomic model is integrated via a biological submodel of the process of farming in sea cages, interrelated with an economic submodel that quantifies the process to consider the economic implications of any change in the farming and market parameters. The purpose of the bioeconomic analysis is to find a harvesting time that maximises the present operational profits in a given time horizon [1]. The model includes three essential factors that influence fish growth: fish weight, water temperature, and feed quantity. In [5] a description of the model is presented. The system has to help to the decision maker to determine the optimal production planning that maximises profits within a finite time horizon, considering the initial time to be the stocking time of the first batch and the final time to be the harvesting time of the last batch. To achieve this objective, the DSS uses a PSO algorithm to identify the optimal production plan. Each particle of the population in the search space represents a sequence of seeding and harvesting operations that generates an economic result. According to the general scheme of all PSO algorithms, each particle has an associated position vector that represents the corresponding production plan. The components of these vectors are the number of days with respect to the previous harvest process, the desired harvest weight of the fish and the fingerling weight in each of the batches performed in the time horizon under the plan.

Two characteristics differentiate a model-driven DSS from the computer support used for a decision analytic or operations research special decision study [11]: (i) A model is made accessible to a non-technical specialist such as a manager through an easy to use interface, and (ii) a specific DSS is intended for some repeated use in the same or a similar decision situation. Models in a model-driven DSS should provide a simplified representation of a situation that is understandable to a decision maker. The goal of making DSS accessible to non-technical specialists implies that the design and capabilities of the user interface are important to the success of the system. In order to provide quick and easy to use access to the system, the aquaculture DSS was implemented using Web-based technologies, with a visual interface that combines HTML language with visual elements developed using Java programming language and MySQL database connectivity (Figure 1). This connection to a database allows determination of the optimal production plan for as many species and farms as have available data. The database stores information about parameters such as different environmental conditions, feeds, market prices, and fingerling prices.

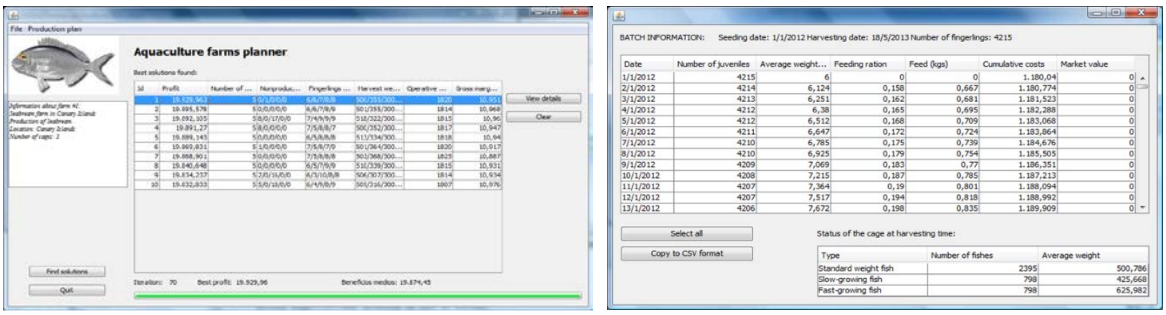


Figure 1: Results windows of the DSS developed after application of optimization process.

3 Application of the aquaculture DSS

As example of practical application, the work shows the application of fish farming DSS to the production of seabream in floating cages in Spain. Firstly, the system is used to support operational decisions. The optimal production plan was obtained for a hypothetical fish farm located in the main production area of this specie in Spain, the Canary Islands. Secondly, DSS is applied to support a strategic decision as site location depending on the environmental conditions. This time, optimal production plan was obtained for the hypothetical fish farm in the Canary Islands and in the Spanish Mediterranean coast (Figure 2). The production was developed under the same cultivation conditions in both locations, with the exception of water temperature.

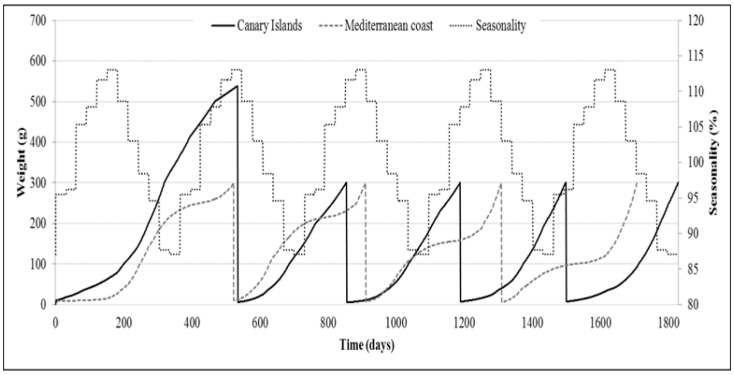


Figure 2: Evolution of seabream weight and the seasonality of its price in the Canary Islands and on the Spanish Mediterranean coast along the five year period simulated.

References

[1] Asche, F., & Bjørndal, T. The economics of salmon aquaculture. Blackwell Publishing Ltd, 2011.

[2] Bjørndal, T., Lane, D. E., & Weintraub, A. Operational research models and the management of fisheries and aquaculture: A review. *Eur. J. of Operational Research*, 156(3): 533-540, 2004.

[3] Bolte, J., Nath, S., & Ernst, D. Development of decision support tools for aquaculture: the pond experience. *Aquacultural Engineering*, 23(1): 103-119, 2000.

[4] Bourke, G., Stagnitti, F., & Mitchell, B. A decision support system for aquaculture research and management. *Aquacultural Engineering*, 12(2): 111-123, 1993.

[5] Cobo, A., Llorente, I. & Luna, L. Swarm Intelligence in Optimal Management of Aquaculture Farms. In Lluis M. Pla-Aragones, editor, *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, Springer-Verlag, 2015.

[6] Conte, F.S. & Ahmadi, A. AQUARIUS: A decision support system for aquaculture. *Proceedings of the Conference 21st Century Watershed Technology: Improving Water Quality and Environment*, Limon, Costa Rica: ASABE, 10-15, 2010.

[7] Ernst, D.H., Bolte, J.P., & Nash, S.S. AquaFarm: simulation and decision support for aquaculture faculty design and management planning. *Aquacultural Engineering*: 23(1-3), 121-179, 2000.

[8] Halide, H., Stigebrandt, A., Rehbein, M.A., & McKinnon, A.D. Developing a decision support system for sustainable cage aquaculture. *Environmental Modelling and Software*: 24(6), 694-702, 2009.

[9] Hargrave, B. A traffic light decision system for marine finfish aquaculture siting. *Ocean and Coastal Management*: 45(4), 215-235, 2002.

[10] Lannan, J.E. Users guide to PONDCLASS: Guidelines for fertilizing aquaculture ponds. Pond Dynamics / Aquaculture Collaborative Research Support Program, Corvallis, OR: Oregon State University, 60, 1993.

[11] Power, D.J., & Sharda, R. Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*: 43(3), 1044-1061, 2007.

[12] Schulstad, G. Design of a computerized decision support system for hatchery production management. *Aquacultural Engineering*: 16(1), 7-25, 1997.

Implementation of Robust Decision Making in Agriculture Planning Decisions using Cloud Computing.

Xavier Ignacio Gonzalez[1]

[1] Facultad de Ingeniería de la Universidad de Buenos Aires
Av. Las Heras 2214, CABA, Rep. Argentina
xgonzalez@fi.uba.ar

Abstract

This document describes a robust decision methodology implemented in the cloud to support farm planning decisions. A model that combines big datasets of prices and simulated yields is built to select robust strategies giving farmers support to mitigate the risk of an unsatisfactory outcome. Additionally, the methodology provides insightful information about those risks.

1 Introduction

Nowadays, it is unarguably that agriculture is one of the industries with the largest potential for development with the rise of Big Data [1]. New sources of diverse data –e.x. customized historical climatic data from IOT sensors, multiples GIS files with geo-located yields extracted from precision agriculture devices, and cheap satellite images– are now accessible to be analyzed by parallel computing in the cloud using several machine learning libraries. An average farmer, who until a few years ago used to select crop allocation with low technicity, now has all these new sources of information to improve the outcome of her decision.

The project “Big Data in Robust Decision Making in Agriculture” –see Acknowledgement section below– is aimed to develop a framework that supports farmers in the decision process, handling large sources of data, analyzing data in the cloud, suggesting good alternatives for land use, and providing additional information to help them deal with deep uncertainty about climate change and future crop price levels.

As part of the project, this study details a decision model within a robust decision-making methodology (RDM). Additionally, more results obtained applying other models and methods like adaptive strategies, Info-Gap Decision Theory, Fuzzy Logic, and Multi-Objective Robust Decision Making (MORDM) will be mentioned in the conference and are not discussed in this abstract. Also, the author expects to show during the presentation a web application where a farmer can test the decision methodology presented in this paper.

2 Methodology: Robust Decision Making

Basically, RDM, introduced in [2], characterizes uncertainty with multiple future views or scenarios. It can also incorporate probabilistic information, but rejects the view that a single joint probability distribution represents the best description of a deeply uncertain future. Second, RDM uses a robustness rather than an optimality criterion to assess alternative policies. There exist several definitions of robustness, but all incorporate some type of satisfying criterion. For instance, a robust strategy can be defined as one that performs reasonably well compared to the alternatives across a wide range of plausible future scenarios. Third, the RDM explores the vulnerabilities of candidate strategies by a supervised machine learning algorithm providing additional insightful information to the decision maker. That is, instead of just suggesting a recommended mix of crops or varieties, it advises decision makers about the situations in which those recommendations will not perform acceptable and extends the recommendation to some feasible alternatives that mitigate those risks. Following section describes how the RDM is adapted for the crop mix allocation decision in farms in Argentine Pampas.

3 Decision Model: farms in Argentine Pampas

3.1 Strategies and Scenarios

A farmer has a limited set of alternatives crops to assign to her available land. A total of 6 different crop managements are considered in this study: 2 for Soybean, labeled as 'Soy1' and 'Soy2'; 2 for Maize; and 2 for the combined Wheat-Soybean within the same season. One possible strategy to choose will consist in a proportion of land assigned to one of those 6 alternatives. For example, 10% of land assigned to 'Soy1' and 90% of land assigned to 'Ma2'. Considering a farm with a resolution of 10% to assign to a different crop management, the total of possible strategies rises to 3003.

To incorporate the uncertainty about the future, multiple scenarios integrate climate and prices information based on historical data. We independently combined 77 climate scenarios and 27 scenarios with plausible crop price levels to obtain a total of 2079 scenarios.

We estimated a Farm Wide Net Margin ($FWNM_{ij}$) for each strategy i and scenario j as a function of simulated crop yields –obtained by DSSAT with climate data– and the simulated crop prices estimated for each scenario. To evaluate each strategy among scenarios, we calculated a metric of robustness called Regret.

3.2 Regret to evaluate Strategies among Scenarios

A relative definition of robustness is often preferable to one based directly on the absolute performance of a strategy. The regret of alternative strategies provides a conceptually and computationally convenient means to help identify robust land allocations and their vulnerabilities [2].

An index of the regret associated with strategy i and scenario j can be computed as the difference between the maximum simulated farm-wide gross margin (R_{ij}) considering all possible strategies for scenario j –that is, the performance of the best possible strategy for this scenario– and the farm-wide gross margin of strategy i in scenario j.

Then, a candidate robust strategy is selected as the strategy that performs well under a broad range of scenarios. In some cases, this initial strategy may be suggested by decision-makers, for example, the recommended crop rotation for a given location. In this study, we set the candidate as the strategy with the lowest value of the third quartile ($Q3_i$, or the 75th percentile) of the regret value distribution for each strategy over the scenarios. The strategy obtained is the crop mix of 50% Soy2, 50% WS1.

Using a decision tree classifier, we explored the candidate vulnerabilities.

3.3 Identification of vulnerabilities in the initial candidate robust strategy

We characterized the multi-dimensional conditions in which the initial candidate strategy is not likely to perform well. In turn, this will allow us to identify alternative strategies that may have a better performance in scenarios where candidate strategy is vulnerable. This stage is often referred to in the literature as "scenario discovery" [3].

An advantage of the decision tree approach is that the sets of conditions that lead to a strategy's bad performance can be easily interpreted and communicated to decision-makers. The tree successively divides the input space with the goal of creating multiple regions that contain outputs of a single class [4]. Each identified portion of space corresponds to a terminal node (labeled "BAD" or "GOOD") in the tree in Figure 1. The predictors used to train the classifier include: (a-c) the prices of maize, soybean and wheat, (d) the ratio of soybean to maize prices, (e) the ratio of soybean to wheat prices, and (f) quartiles of the historical distribution of total rainfall between September of one calendar year and March of the following year. Going down along the branches of the tree, the splits show that the outcome of strategy 931 strongly depends on scenario parameters such as the ratio of soybean to wheat prices, the maize price, and the quartile of rainfalls.

4 Results

For each strategy, we compute the third quartile (Q3) of regret values separately for "good" and "bad" scenarios, as classified by the tree in Figure 1. Then we use these quantities as coordinates for a plot of a subset of strategies with low regret values showed in Figure 2.

Initial strategy 931 had the lowest Q3 of regret in all actual scenarios. To mitigate the risks identified by the tree, more robust alternatives can be found. An “ideal” alternative to strategy 931 would have low regret values under both good and bad scenarios. Such strategy would be preferred over all other possible land allocations, as it would perform satisfactorily regardless of the scenario that might occur. Unfortunately, Figure 2 does not show an ideal strategy, which would be in the lower left of the figure. Instead, we define a “low-regret frontier” defined by the lower edge of the cloud of strategies; this frontier is indicated by a dashed line in Figure 2. Strategies along the low-regret frontier dominate other strategies: for a given regret in “good” scenarios (x-axis), strategies above the frontier have higher (i.e., worse) regret; the same situation happens with respect to the y-axis.

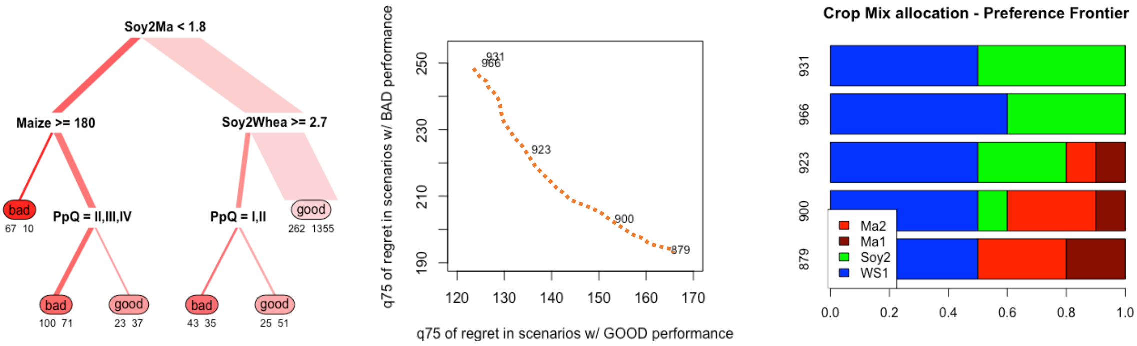


Figure 1 (left): Classification tree predicting “bad” or “good” performance for candidate strategy. The numbers in each terminal node indicate respectively the number of “bad” and “good” records in that node (note that the terminal nodes are not 100% of one class or the other, i.e., they are impure). Figure 2 (middle): Calculated 3rd quartile of regret for each land allocation strategy in both “good” and “bad” scenarios (defined regarding predicted performance of candidate strategy 931). Figure 3 (right): Land allocations for the strategies in the preference frontier.

5 Implementation

The instance of the model described in this study requires the evaluation of all 3003 strategies under each of the 2079 scenarios, which results in more than 6 million records. This number evidences the need of a scalable system to easily include more crop managements and multiple farms simultaneously. Therefore, the script coded in R was implemented on Microsoft Azure cloud computing services platform.

6 Acknowledgment

This research was funded by the Peruilh PhD scholarship and the Argentine National Agency for Scientific and Technological Promotion under the project code PICT134.

References

[1] Wolfert, Sjaak, et al. "Big Data in Smart Farming–A review." Agricultural Systems 153 (2017): 69-80.

[2] Lempert, Robert J., et al. "A general, analytic method for generating robust strategies and narrative scenarios." Management science 52.4 (2006): 514-528.

[3] Bryant, Benjamin P., and Robert J. Lempert. "Thinking inside the box: a participatory, computer-assisted approach to scenario discovery." Technological Forecasting and Social Change 77.1 (2010): 34-49.

[4] Therneau, Terry M., and Elizabeth J. Atkinson. An introduction to recursive partitioning using the RPART routines. Vol. 61. Mayo Foundation: Technical report, 1997.

Decision support system for farmland fertilization based on linear optimization with fuzzy cost

Esmelin Niquin-Alayo[1], Edmundo Vergara-Moreno[2], Marks Calderón-Niquín[3]

[1]National University of Santiago Antunez de Mayo
Av. Centenario s/n, Huaraz, Perú.
eniquin@unasam.edu.pe

[2] National University of Trujillo
Av. Juan Pablo II s/n, Trujillo, Perú
evergara@unitru.edu.pe

[3] University ESAN
Av. Alonso de Molina 1652, Surco, Lima, Perú
mcalderon@esan.edu.pe

Abstract

In this work, the fuzzy sets theory for the mathematical modeling of fertilization farmland problems, considering fuzzy sets, is used. With the support of programming languages and computer tools, a software called FERTIDIF to contribute in making decisions about the utilization of fertilizer and nutrients required for a particular crop, has been designed. The problem of fuzzy fertilization (PFD) has been solved by means of the adaptation of methodologies of solution proposed by Lai-Hwang and Leberling; methodologies transforming a fuzzy problem with fuzzy costs into a multi-objective optimization problem and solved easily by using the FERTIDIF software.

1 Introduction

Agricultural productivity depends on several factors but primarily on the fertility of the soils. The soil is fertile when it has enough quantities of nutrients to the plants. In order that vegetables can capture the nutrients in optimal quantities, the soil must exhibit ideal physical fertility (texture, porosity, permeability and depth appropriate to favor the circulation of air and water), chemical fertility (with appropriated pH, CEC(Cation exchange capacity), EC (electric conductivity), P, N and K) and ideal biological fertility (organic matter, drainage, proper agrochemical products) [2].

To support the optimality of the soil is necessary to fertilize, adding the necessary nutrients to obtain a high level of productivity. In the planning of the fertilization, traditional tools, classic models of optimization, expert systems of decision or heuristic are in use [5]; but only very few use methods of fuzzy optimization.

Agricultural production and its fertilization is an activity of several periods depending on the type of cultivation. For example, the cultivation period of potato is between 3.5 to 6.5 months; according to the variety, the sugar cane varies from 14 to 17 months in the first cut and from 11 to 13 months in the following cuts, etc. In the planning of expenses for activities with longer execution time, it is advisable the use of fuzzy models [3, 9, 11, 12].

For these reasons, in this work we build the model with fuzzy cost with the aim to determine the optimum fertilization plan of a crop, and to resolve it the software FERTIDIF has been developed.

2 Model and method of solution to the problem of fertilization

2.1 Model of fertilization

We use the diet model, also established, solved and applied to the poultry farms in Peru [12] and to the farms of cattle in Argentina [11]. Here, it is formulated with fuzzy cost:

$$\text{Min} \sum_{j=1}^{n} \tilde{c}_j x_j$$
Subjec to:
$$p_i \leq \Sigma_{j=1,\dots,n} a_{ij} x_j \leq P_i, \quad i = 1, \dots, m \tag{1}$$
$$m_j \leq x_j \leq M_j, \qquad j = 1, \dots, n$$

where: x_j represents the quantity of fertilizer j to be included in fertilization, a_{ij} represents the quantity of nutrient i contained in fertilizant j, the m first constraints limit the total quantities of nutrients in fertilización, that is, every nutrient i must not be less than the quantity p_i or exceding the quantity P_i (necessary minium and maximum), while the other n constraints limit the quantity of every allowed fertilizant j, not to be less to m_j or exceeding M_j, $\tilde{c}_j$: the unitary cost of fertilizer j, in fuzzy version, represented by a triangular fuzzy number [1, 4, 8].

2.2 Method of solution to the model of fertilization

There are several methods of solution to the model (1) considered as a linear optimization problem with fuzzy costs [3, 6, 7, 10]. Because of the simplicity to implement, the method of Lai and Hwang, transforming the objective function of fuzzy costs in multiobjective, has been used [6]:

$$\min \; z = (-(c^m - c^p)^T x, \; (c^m)^T x, (c^o - c^m)^T x), x \in F \tag{2}$$

where: F is the set of values x satisfying the constraints of model (1).
To the solution of model (2) two methods, one proposed by Zimmermann [13] and improved by Lai and Hwang [6], and other one proposed by Leberling [7], have been used.

Then, from the fuzzy formulation, by using the best decision criterion [6], we obtain the following classical linear model:

$$\text{Max } \alpha; \; x \in F, \; \mu_i(x) \geq \alpha \;, \quad i = 1,2,3 \tag{3}$$

Solved by using the simplex method [9].

3 Development of software FERTIDIF

3.1 The interface

The main menu has the typical options: File, Run, Report and Help; and special options of Fertilizers and crops. In the Run menu we find the options of Lai-Hwang y Leberling that are the present methods of solution in the software. The menu Fertilizers includes the option to modify the data of these both in their composition as well as their fuzzy costs.

3.2 Data Base

When FERTIDIF is used for the planning of agricultural fertilization, there is a need for a data series of the ground where we plan to grow a particular type of crop, which will require the fertilization to achieve a high productivity. In this sense, the data of the fertilizers and manure, as well as crops, have been included in the software.

3.2.1 Fertilizers and Manures

Includes the relationship of the existing fertilizers, with their nutritional components and their fuzzy prices expressed by means of a shortlist. The modifications of this base are done through the menu Fertilizers or Manures. At the time of using the system, we select the Fertilizers or Manures available in the market and whose data are stored.

3.2.2 Crops

When entering in this option, the data entry window of the ground, type of ground, absorption capacity in relation to the type of crop, both the nutrients as well as the fertilizer and manure, is activated to enter the system. Here the quantity of every type of nutrient required for every crop is included.

4 Application of software in Agricultural planning

As an illustration, the fertilization of a hectare of lands in the Peruvian coast for the cultivation of potato is planned. The FERTIDIF serves for the preparation of fertilization plans in real dimensions; however, in this illustration only three basic nutritional substances are considered: nitrogen (N), phosphorus (P_2O_5 just represented by P) and potassium (K_2O, just K). The main data are provided in Table 1.

Table 1: Fertilizers with components and prices in the market

Fertilizer	% N	% P_2O_5	% K_2O	Requeriment Min-Max	Cost Kg/oles
Super Guano	20	20	15	150-190	$\tilde{c}_1$=(1; 1,2;1,5)
Guano Islands	10	10	2	120-160	$\tilde{c}_2$= (0,8; 1; 1,2)
Ammonium Phosphate	16	48	0	115-155	$\tilde{c}_3$=(1; 1,15; 1,3)

The solution by using FERTIDIF, provides α =0,616279; Minimum: $\tilde{z}$=(776,274; 931,064; 1162,55) (Lai-Hwang) and α =0.5, Minimum : $\tilde{z}$=(769,307; 923,168; 1153,96) (Leberling).

Acknowledgment: A Part of this work has been developed within the framework of the CYTED Project P515RT0123 (BIGDSSAGRO): Ibero-American Network of Agro-Big Data and Decision Support Systems (DSS) for a sustainable agricultural sector.

References

[1] Brunelli, M.; Mezei, J. How different are ranking methods for fuzzy numbers?: A numerical study. International Journal of Approximate Reasoning 54: 627-639, 2013.

[2] Echevarría, H. E.; García, F. O. Fertilidad de suelos y fertilización de cultivos. Ediciones INTA, Buenos Aires, Argentina, 2005.

[3] Ezzati, R.; Khorram, E.; Enayati, R. A new algorithm to solve fully fuzzy linear programming problems using the MOLP problem. Applied Mathematical Modelling 39(12): 3183-3193, 2015.

[4] Ezzati, R.; Allahviranloo, T.; Khezerloo, S.; Khezerloo, M.An approach for ranking of fuzzy numbers. Expert Systems with Applications 39(1): 690-695, 2012.

[5] Hernández, J.C. Edafología y fertilidad. Universidad Nacional Abierta y a Distancia. Bogotá, Colombia, 2013.

[6] Lai, Y.J.; Hwang, C.L. A new approach to some possibilistic linear programming problems. Fuzzy Sets and Systems 49(2): 121-133, 1992.

[7] Leberling, H. On finding compromise solutions in multicriterial problems using the fuzzy min operator. Fuzzy Sets and Systems 6(2): 105-118, 1981.

[8] Liang, D.; Liu, D.; Pedrycz, W.; Hu, P. Triangular fuzzy decision theoretic rough sets. International Journal of Approximate Reasoning 54(8): 1087-1106, 2013.

[9] Luenberger, D.; Ye, Y. Linear and Nonlinear Programming. . International Series in Operations Research & Management Science 228, Springer US, New York, USA, 2016.

[10] Luhandjula, M. K. Fuzzy optimization. Fuzzy Sets and Systems 274 (C): 4-11, 2015.

[11] Pelta, D.A.; Verdegay, J.L.; Cadenas, J.M. Introducing SACRA: A Decision support. Applied Decision Support with Soft Computing 124: 391-401, 2012.

[12] Vergara-Moreno, E.; Rodríguez-Novoa, F.; Saavedra, H. Métodos de optimización lineal difusa para la planificación nutricional en granjas avícolas. Mosáico Científico 3: 16-29, 2006.

[13] Zimmermann, H. J. Fuzzy programming and linear programming with several objective functions, Fuzzy Sets and Systems 1(1): 45-55, 1978.

Assessing traceability system adopted by the Mango supply chain in Colombia: An analysis of the asynchrony in the inventory and food quality

Milton Herrera[1], Javier Orjuela[2]

[1] Universidad Piloto de Colombia
milton-herrera@upc.edu.co

[2] Universidad Distrital Francisco José de Caldas
jorjuela@udistrital.edu.co

Abstract

The skill of finding the quality problems of a product along process is associated with the performance and efficiency of the traceability system adopted by supply chain. For that reason, technological changes in the traceability system involve opportunities and drawbacks, which must be analyzed in terms of supply chain performance. In this sense, this article presents a simulation model that evaluates the implementation of traceability technologies through a case study for Mango supply chain. Results evaluate the asynchrony between traceability systems of radio frequency identification and barcode along supply chain as well as their effects on inventory and food quality.

1 Introduction

Technological management of traceability systems has potential impacts on the performance of food supply chain. Previous studies showed the economic impacts of RFID technology on sales and revenue of retail operations [1] as well as its diffusion and adoption on organizations [2]. Likewise, the technological changes of adopted traceability technology may affect the food quality and inventory control. Considering that technological changes are continuality creating new opportunities for product development and industrial diversification, these opportunities need to be analyzed through a better understand of technology management as a dynamic capacity [3]. Despite that the technological growth has led to the implementation and acquisition of technologies that improve the performance of supply chain, the technology changes have been affected the asynchrony in the operations of supply chain [4]. In this sense, the policies of technology management in terms of investment and selection technological play a role important for food supply chain.

Technology adoption of traceability requires appropriate technology management policies for implementation, which involves an analysis of long-term dynamics. The traditional approaches to evaluating adopted technologies show high and simplifying level or a very detailed and specific level [5]. In this case, the system dynamics simulation may develop insights and learning with respect to the dynamic behavior along supply chain as well as better understand the adoption of technologies. In addition this, the simulation models with system dynamics are particularly useful for analyzing scenarios and technology management policies associated to complex causality. Several studies reported on a methodology based on system dynamics for technological assessment, which is showing the implication of the technological adoption in the long term [1], [2], [5]–[8]. However, these studies not take into account the performance of supply chain as well as its relation with the traceability system.

In supply chain management, a traceability system has a great important due it is considered an effective means of finding and resolving quality problems [9]. Like this, the tracking technology in the food supply chain adopted can improve the quality control of products and setup time associated with the process. The aim of this article is evaluated the traceability systems in terms of inventories and quality control for the Mango supply chain. The novelty of this research is the dynamic perspective that proposes for the evaluation of traceability technology and its effects in the inventory of perishable food in supply chain.

2 Model description

2.1 Simulation model

The simulation model is represented by stock and flow diagram. This model presents the traceability system and quality control of food supply chain. The stock and flow diagram includes three subsystems to assess the adoption of traceability system and quality control in the Mango supply chain. The first subsystem is related with the effects of quality improvement programs on the adoption of traceability system, which generate an impact on quality control of the food. Equations (1) and (2) calculate the quality shortfall and the percentage of quality product controlled by the capacity of *traceability system* (tsa), respectively. Second subsystem represents the effects of the food demand on the technologies investment of production and traceability. The *expected demand of food* (EDF) is calculated as observed in Equation (3) and (4). The last subsystem includes the impact on the product quality generate by the estimated life cycle of the Mango with regard to the amount of product on the inventory. The main equations describing these subsystems are here exposed:

$$Quality\ shortfall = desired\ product\ quality - product\ quality\ \ [\%] \quad (1)$$

$$Product\ quality = tsa * effect\ \ of\ life\ cycle\ \ [\%] \quad (2)$$

$$EDF(t) = \ EDF(t - dt) + \int_{t=o}^{T} [\, Change\ of\ demand(s)]ds \qquad [Tn] \quad (3)$$

$$Change\ of\ demand = EDF * product\ quality\ \ [Tn/Year] \quad (4)$$

The base simulation experiment was done based on assumed initial simulation parameters. A synthesis of the main assumptions for the simulation model as well as some input data are presented:

- Given the dynamic of change in the traceability systems evaluated by [2], the simulation horizon was set at 25 years due to time that using for development new technologies.
- The initial parameters are associated with the Mango demand in Colombia reported by the Ministry of Agricultural. In this sense, all the data is indexed to the base year, 2015.
- For the estimation of inventories, the study elaborated by [7] is used in the model. According to study by [10], the percentages of traceability system was considered.

3 Simulation results

The proposed model has been simulated from 2017 to 2042 for better understand the asynchrony in terms of inventories, traceability system and quality shortfall along the Mango supply chain. The results of asymmetries among producer, agribusiness, wholesale and retail trader related with the inventories are shown in Figure 1. This behavior of asynchronies in the inventories of the supply chain was addressed and analyzed by [6] through a simulation model. That study presented the effects of time delays in the shipments on inventories among the stakeholders of supply chain. In our simulation model, the asynchronies in the Mango supply chain affect the quality control due the delays in the adoption of different traceability systems. The results of simulation show that there is a considerable difference between producer inventories and retail trader, which generate difficulties the quality control of traceability system.

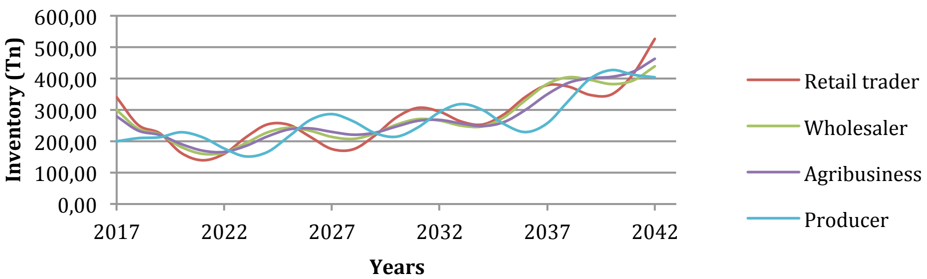


Figure 1. Asynchrony of the inventories in the Mango supply chain

When exist different traceability technologies along supply chain can produce an effect of asynchronies on inventories and quality controls. This condition has an impact that affects the performance of supply chain (seasonal inventory average and product quality). Different experiments were considered assessing some possible combinations for the traceability systems adopted by supply chain. The experiments present the synchronies in the supply chain through the performance measure associated with seasonal inventory and product quality. The results show that a better performance in terms of inventory and product quality is base on traceability technologies homogeneous with 258,15 ton seasonal inventory average and 97,14% product quality average along supply chain of Mango. Experiments validate that a common selection of traceability technologies is associated with a better performance in the product quality average. Therefore, the model estimates synchronies as from similarly traceability technology in each players of the supply chain.

4 Conclusions

The relationship of the actors in the supply chain fruit is part of different dynamics that regulate their behavior. In this sense, the dynamics of technology for traceability in the supply chain has a growth that generates impacts on improving the flow of material and information. Therefore, the decision to change and implementation of traceability technologies in the food supply chain requires comprehensive analysis oriented models of relationships and flows between actors of the chain.

References

[1] A. De Marco, A. C. Cagliano, M. L. Nervo, and C. Rafele, "Using System Dynamics to assess the impact of RFID technology on retail operations," *Int. J. Prod. Econ.*, vol. 135, no. 1, pp. 333–344, 2012.

[2] Y. Chen, "Understanding Technology Adoption through System Dynamics Approach: A Case Study of RFID Technology," *Embed. Ubiquitous Comput. (EUC), 2011 IFIP 9th Int. Conf.*, pp. 366–371, 2011.

[3] D. Cetindamar, R. Phaal, and D. Probert, "Understanding technology management as a dynamic capability: A framework for technology management activities," *Technovation*, vol. 29, no. 4, pp. 237–246, 2009.

[4] J. Orjuela-Castro, M. Herrera-Ramirez, and W. Adarme-Jaimes, "Warehousing and transportation logistics of mango in Colombia : A system dynamics model," *Rev. Fac. Ing.*, vol. 26, no. 44, pp. 71–85, 2017.

[5] E. . Wolstenholme, "The use of system dynamics as a tool for intermediate level technology evaluation: three case studies," *J. Eng. Technol. Manag.*, vol. 20, no. 3, pp. 193–204, 2003.

[6] J. D. Sterman, *Business dynamics: Systems Thinking and Modeling for a Complex World*, no. December 1999. McGraw-Hill, 2000.

[7] M. M. Herrera and J. Orjuela, "Modelo para la implementación de tecnología de trazabilidad RFID en la cadena de suministro frut ícola en las operaciones de picking bajo un enfoque integral y dinámico difuso," Universidad Distrital Francisco José de Caldas, 2014.

[8] T. Stavredes, "A system dynamics evaluation model and methodology for instructional technology support," *Comput. Human Behav.*, vol. 17, no. 4, pp. 409–419, 2001.

[9] K. Zhang, Y. Chai, S. X. Yang, and D. Weng, "Pre-warning analysis and application in traceability systems for food production supply chains," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2500–2507, 2011.

[10] M. M. Herrera, M. Becerra, O. Romero, and J. A. Orjuela Castro, "Using System Dynamics and Fuzzy Logic To Assess the Implementation Rfid Technology," in *32st International Conference of System Dynamics Society*, 2014.

A Simulation Model to Analyze the Payback Period of a Sow Farm Using the Transient State

Marco A. Montufar[1,2,3], David Fernando Muñoz [4]

[1] Instituto Tecnológico y de Estudios Superiores de Monterrey, Toluca, Estado de México, C.P. 50110
[2] Departamento de Matemática, Universidad de Lleida, 73, Jaume II, 25001, España
[3] Universidad Autónoma del Estado de Hidalgo, Pachuca, Estado de Hidalgo, C.P. 42184, México.
montufar@uaeh.edu.mx
[4] Instituto Tecnológico Autónomo de México,
Río Hondo # 1, C.P. 010080 México DF, México
davidm@itam.edu

.Abstract

This work presents the development and analysis of a discrete simulation model to study the payback of the investment in a pig farm dedicated to the production of piglets. The model considers the stages of mating, gestation, and lactation typical of the process. The random variables considered were litter size, success at fertilization, and residence times at each of the stages mentioned above. The results show the risks of reaching a certain return period, when the sows remain a certain number of cycles on the farm.

1 Introduction

It is common for investors to be concerned about profitability, as well as for the prompt recovery of money invested in a business. Engineering Economy gives us a rational way of measuring the profitability and liquidity of investments. Performance measures to measure profitability are for example: the net present value, the internal rate of return, the annual value, etc..; In the same way the simple and discounted payback are the traditional ways of measuring liquidity. [1].
The payback of an investment is the speed with which money is recovered, and has been defined as the number of periods necessary for an annual income to equal the initial investment. Under this simple definition are assumed some aspects that many times are not well understood. The first point is that the analyst must somehow estimate the inflows and outflows of the business through the useful life, and then calculate the annual equivalent discounted. In many textbooks that value is given as a data, but the question that arises is: how was calculated? It is probable that this annual income will be positive from the first analysis period, and then the recovery period will be calculated by a simple division between the initial investment and the annual positive flow. Many of these average cash flow are based on long-term behavior, so a new question arises: are long-term estimates valid to be representative of short-term behavior?
The production of piglets on farms has been studied from the stochastic point of view [2,3,4], with simulation models and Markovian decision processes. Some of these models have focused on determining the proper cycle of replacement of the mother sows, while others on the optimal determination of some decision variable, such as the moment of weaning.
This study aims to measure the risk of reaching a certain payback, an important element in the decision-making of a farmer, through the study of the transient state of the system. To this end, we propose a simulation model that considers the main stages of the production process of the farms

2 Materials and Methods

The model presented here is based on the previous work of [5], which consider the stages of fecundation, gestation and lactation as main processes (see Fig. 1), in these studies the objective was to determine the optimal cycle of permanence of the mother sows in order to achieve maximum daily profit. These works take into account the steady state of the system, since the interest is focused on profitability, rather than on the liquids of the business. Now we model this system to study the payback for a farm of 120 mother sows. An important random variable is the number of piglets born and suitable for sale, the main factor of income in the company, Figure 2 shows the box diagram for this variable with respect to the parity number. It would seem obvious to observe this graph that the optimal number to conserve a sow would be 4 or 5 cycles, stages in which the prolificacy is higher, however, the dynamics of the system itself makes that number be eight, as demonstrated in [5], the reason is that replacing it in the cycle 4 or 5, would have the main drawback of introducing to the farm a new sow less prolific than a sow in cycle eight, and would not take advantage of its capacity to generate piglets still in the cycles 6 to 8.

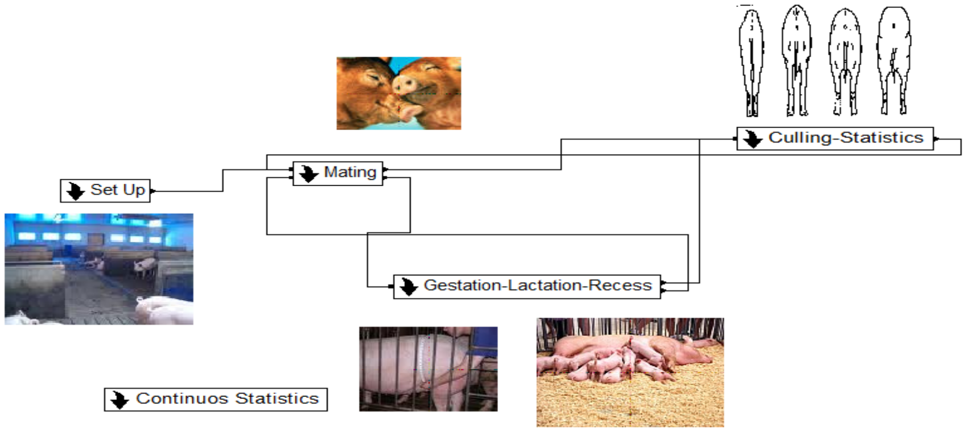


Fig. 1 General view of the simulation model

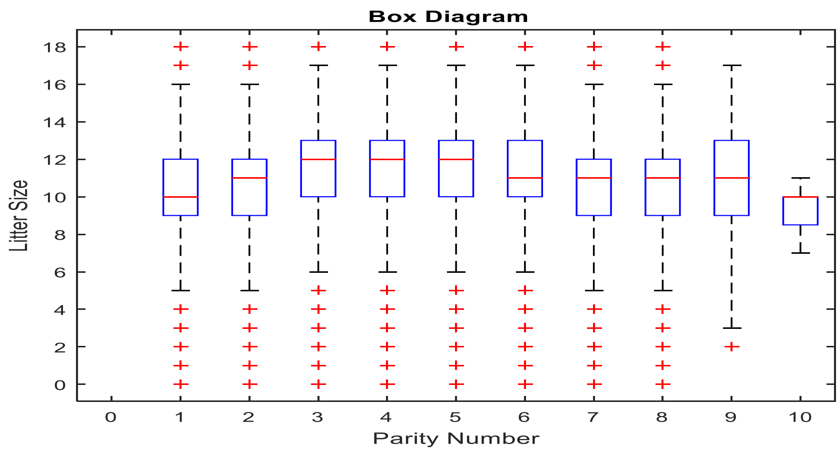


Fig.2 Litter size variability

Our model makes use of the definition of payback established in [1], which is the minimum value that satisfies the following inequality.

$$\sum_{k=1}^{j}(R_k - E_k)(P/F, i\%, k) - I \geq 0 \tag{1}$$

Where, $R_k y\ E_k$, are the income and expenses in the period k respectively, $i\%$ is the MARR, and I is the initial investment generally realized at the beginning of period one.

3 Results and Conclusions

After running the model a certain replication number, to have a 95% of confidence level, the realization of a typical curve representative of the expression (1) was observed, for example, in the case of preserving the sows nine cycles, a replica showed the behavior in Fig. 3. Here we can see how the net present value equal to zero, which led us to consider two values for payback in expression (1), the minimum and maximum value reached. Figure 4. shows the level of risk to reach a return period between 2 and 3 years.

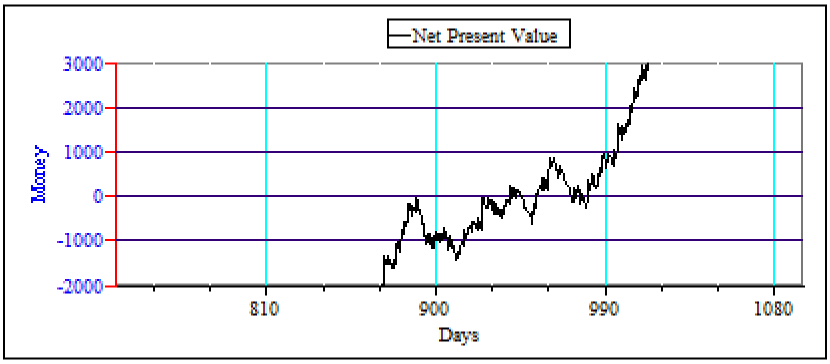


Fig.3 Realization of the NPV

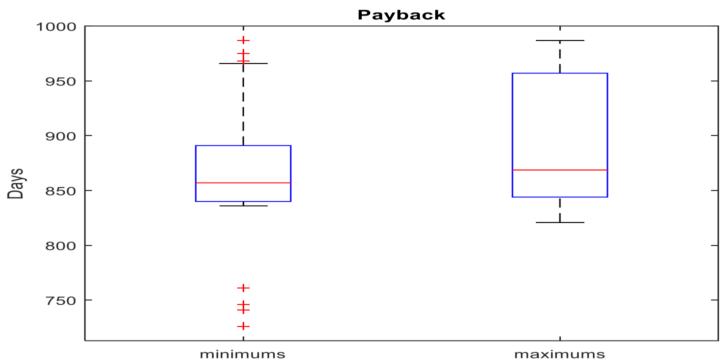


Fig.4 Payback variability

Marco A. Montufar. wish to acknowledge CYTED program to support the thematic network BigDSS-Agro (P515RT0123).

References

[1] Sullivan William G., Wicks Elin M., Koelling C. Patrick, Engineering Economy, . Sixteenth edition., Pearson, USA (2015).

[2] Martel G., B. Dedieu and J.-Y. Dourmad; *Simulation of sow herd dynamics with emphasis on performance and distribution of periodic task events*, Journal of Agricultural Science : 146, 365–380 (2008).

[3] Plá L.M., J. Faulín. and S.V. Rodríguez; *A linear programming formulation of a semi-Markov model to design pig facilities*, Journal of the Operational Research Society; 60, 619 –625 (2009).

[4] Rodríguez-Sánchez Sara V, Lluís M. Plà-Aragonés , Victor M. Albornoz; *Modeling tactical planning decisions through a linear optimization model in sow farms*, Livestock Science: 143: 162–171 (2012)

[5] Montufar B. Marco A., Luis M. Plá Aragonés, Marco A. Serrato García, y Diego Braña Varela; *Análisis y Simulación de Políticas de Reemplazo en Granjas de Explotación Porcina*; Revista Investigación Operacional vol., 34 , no 2. 128-139, (2013)

Simulation of cattle farms with System Dynamics in a serious videogame. Case: SAMI

Urbano Gómez[1], Oscar Gómez[2]

[1] Universidad Pontificia Bolivariana (UPB)
Km 7 vía Piedecuesta, Edificio K, Oficina 107, Floridablanca, COLOMBIA
urbano.gomez@upb.edu.co

[2] Universidad Pontificia Bolivariana (UPB)
Km 7 vía Piedecuesta, Edificio K, Oficina 107, Floridablanca, COLOMBIA
oscar.gomezs@upb.edu.co

Abstract

SAMI is a videogame for learning about bovine cattle production systems, which has been created by the Research Group on Informatics and Telecommunications (INTELEC) of the Universidad Pontificia Bolivariana at Bucaramanga. The game was developed with Unity Engine and its rules are based on System Dynamics. On each game, the player must take decisions and do actions that allow him to correctly run a farm. Those actions are stored in a website which, in order to ease the learning process, provides feedback by comparing the executed actions with those which would have been ideal.

1 Introduction

SAMI aims to teach its players about how bovine cattle production systems work and behave, by providing a close-to-reality interaction and offering feedback. The game provides players with enough resources and mechanics to play and stay engaged, and at the same time helps them to develop skills in taking decisions about selecting the most appropriate breeds, feeding, milk production, treating diseases, buying and selling, among others.

System Dynamics models support the game mechanics. These models represent a cattle farm that the player must administer by taking decisions to simulate experiences. The models are simulated in steps that correspond to months in the game, and the player receives feedback for his decisions both during and after each month. This feedback allows the player to learn, and at the same time provides a sense of progression that motivates continued effort.

2 Contextual Framework

The following are the main topics comprised by SAMI:

2.1. Cattle production systems

In order to administrate a cattle production system, several important factors must be controlled, regarding the amount of available resources and the status of the different related markets, among other [1]. In addition, it is mandatory to understand the different complexities within the system [2] in order to be able to propose and execute strategies for improving it.

2.2. Simulation with System Dynamics

System Dynamics (SD) is a methodology for learning, explaining and recreating phenomena of interest in terms of simulation models [2], which allow the observation of how can the system behave under different circumstances. It works under the assumption that phenomena can be studied as dynamic systems, which can be explained and understood by making experiments of different situations (scenarios), and which show the existing feedback that may exist among variables of the model.

System Dynamics uses five languages of increasing complexity:

1. System verbalization: Explanation in natural language of the available knowledge about the phenomenon.
2. Causal loop diagram: Represents the system structure depending of variables and the relations between them.

3. Stocks-flows diagram: Representation based on elements of System Dynamics, such as flows, level variables, parameters, auxiliary variables, exogenous variables and delays.
4. Mathematical equations: Linear or non-linear differential equations to evaluate the evolution of a variable in time.
5. Model simulations: Visualizations of the data obtained after running the simulations.

2.3. Serious games

Serious games are those designed not only to entertain, but also to educate its players about some topic [3], while usually similar to simulations, they aim to use fun as a way of enhancing the learning process.

3 Proposal

SAMI simulates several characteristics of the system, including feeding, growth, meat and milk production, breeding, health and death, and takes into account specific features of each breed. Players can easily interact with the system through an interface. The development of the game involved the six elements shown in Figure 1: Of **Cattle Production System** (which, in turn, is composed by five subsystems: demographic, biophysical, productive, financial and health, as expressed in [4]) was realized a **Simulation Model** with the basis for defining the equations that produced a **Program in C#** for implementing the **Videogame in the Unity** framework. The player makes decisions while playing the videogame; those decisions are saved in a XML file, which is exported to a web server after the playing session. The server simulates the results of the game while taking correct decisions; with those results, the player can generate **Reports in the Web Information System** which allow him to compare his choices with better ones, which is expected to allow him **To Learn**.

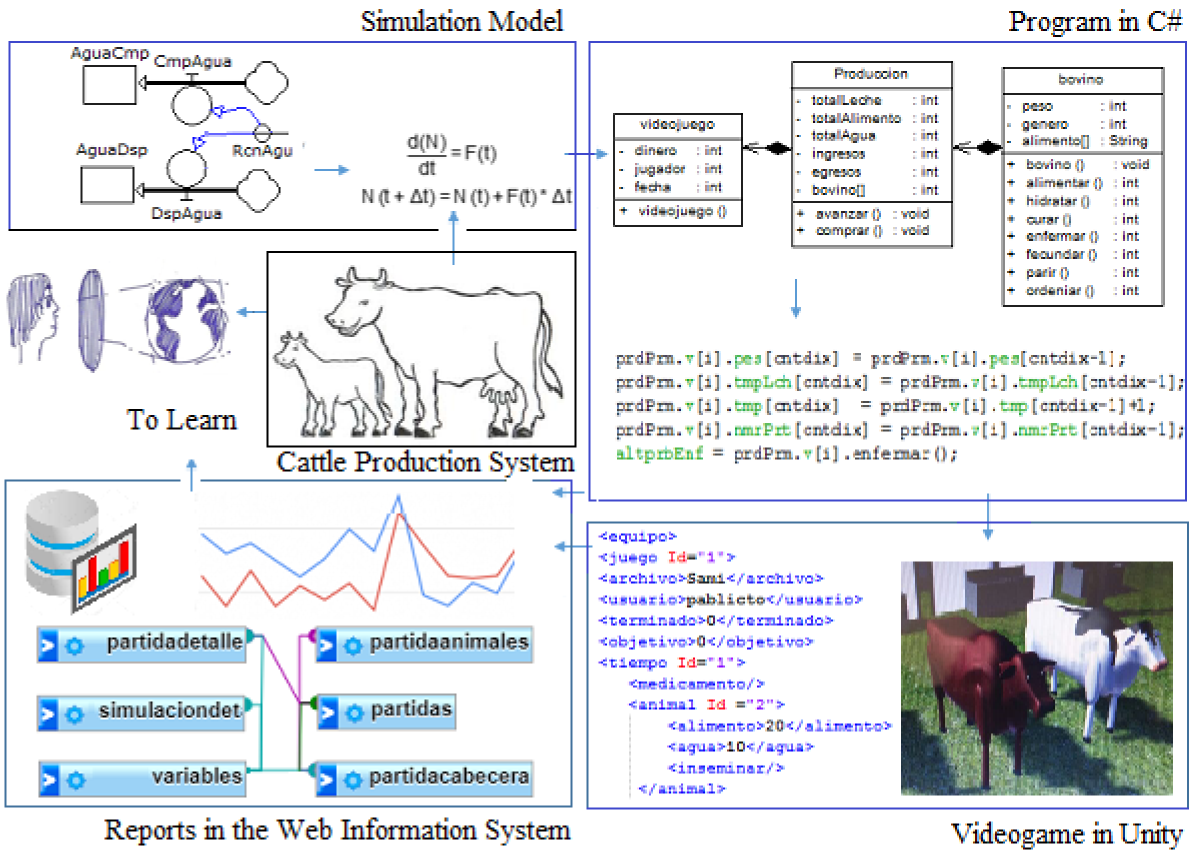


Figure 1. Components of the videogame.

As an example, consider the causal loop diagram and the summary stocks-flows diagram shown in Figures 2a and 2b, respectively. They show six feedback cycles and the main variables of the system, and allow some analysis such as the following:

- Figure 2a shows that empty cows can be bred so they become pregnant cows, which give birth to produce calves and become milking cows. Two cycles are created if calves are left to grow and turn into empty cows, and if milking cows are bred again. At the same time, both calves and milking cows can provide income if sold or milked, Besides it shows that buying empty cows, food or water generates expenses, which decrease the amount of available money.
- Figure 2b shows some of the main variables within the system are the amount of cows, the weight of each cow, the amount of water and food given to each cow and the available money. Stocks (rectangles) in represent them. These values can be changed by flows (circle-shaped valves), which represent events such as income, expenses, purchases, breeding and birth.

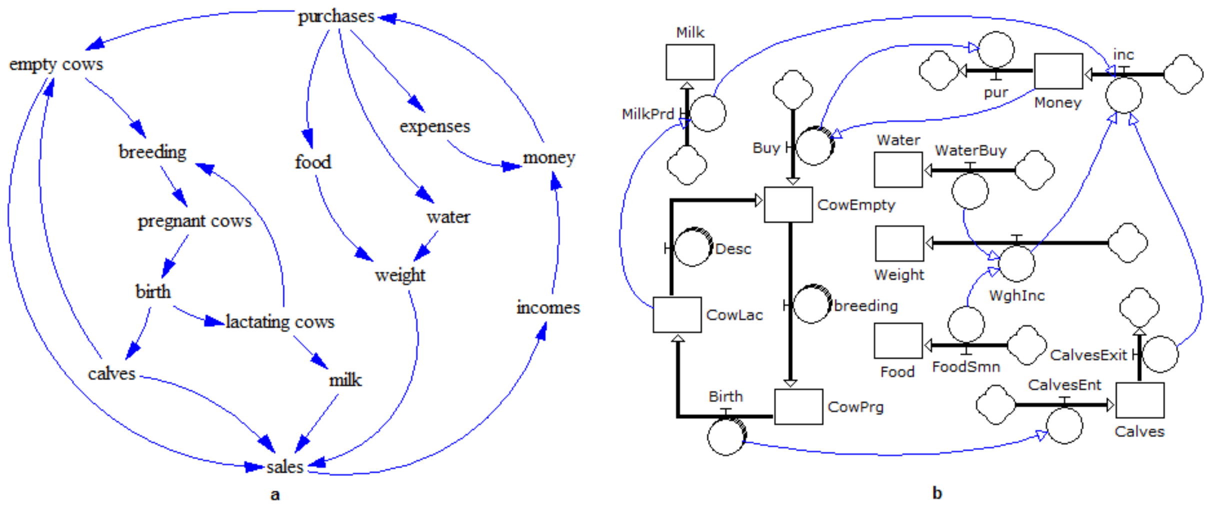


Figure 2. Basic causal loop diagram and stocks-flows diagram

4. Conclusions

Simulation models allow the execution of experiments with variations of their parameters. With enough feedback, these experiences improve the knowledge about systems. In SAMI, some of those parameters are the initial budget, animal breeds and costs of buying and selling supplies; by varying them when playing, the player further understands the administration process of cattle productions systems.

It is possible that during gameplay, players make an unconscious analysis of the dynamic system of SAMI, and that the detection of feedback loops in that system is an important ingredient for achieving a sense of progression, engagement and fun. This hypothesis may also apply to other games, both virtual and physical.

References

[1] C. Phillips, Principios de Producción Bovina, Zaragoza: Acribia S.A, 2003.

[2] H. Andrade, I. Dyner, A. Espinosa, H. López and R. Sotaquirá, Pensamiento Sistémico, Diversidad en búsqueda de unidad, Bucaramanga: UIS, 2001.

[3] Wein, Anne, and Labiosa, William, 2013, Serious Games Experiment toward Agent-based Simulation: U.S. Geological Survey Open File Report 2013-1152, 30p.

[4] U. Gómez Prada, H. Andrade Sosa and C. A. Vásquez, "Lineamientos Metodológicos para construir Ambientes de Aprendizaje en Sistemas Productivos Agropecuarios soportados en Dinámica de Sistemas," Información tecnológica (http://goo.gl/DTJ6df), vol. 4, no. 26, p. 11, 2015.

Forecasting Pesticides Usage Trends Based on Evolutionary Scientific Workflows

Sergio Manuel Serra da Cruz[1], Anderson Oliveira[1],
Raimundo José Macário Costa[1], Fabrício Firmino de Farias[2]

[1] Federal Rural University of Rio de Janeiro
Programa de Pós Graduação em Modelagem Matemátca e Computacional
BR465, Km-7, Prédio Principal – Sala 80 – Seropédica – Rio de Janeiro – Brazil
serra@pet-si.ufrrj.br, anderson@infoassociados.com.br

[2] Federal University of Rio de Janeiro
Programa de Pós Graduação em Informática
Núcleo de Computação Eletrônica – Cidade Universitária – Rio de Janeiro – Brazil
firminodefaria@ufrrj.br

Abstract

Evolutionary computing (EC) relies on the biological behavior of living organisms to obtain better solutions to solve complex problems in polynomial time. Nowadays, the traditional scientific workflow management systems (SWfMS) offer little support for the adoption of EC. Thus, to bridge this gap, we propose the *VisPyGMO* which uses the Island and Archipelago Model, a traditional EC paradigm to extend the functionalities of the SWfMS by incorporating easy-to-use reusable modules of EC algorithms. Besides, we discuss a real use case using such algorithms in VisTrails; we developed evolutionary scientific workflows which analyzed more than 20 years of historical data to predict the future use of pesticides in the agribusiness.

1 Introduction

Big data, decision-making systems, operational research and Evolutionary Computing (EC) are likely to be of tremendous benefit to the agricultural sector and related agri-food industries of many countries [1]. Perhaps one of the greatest advantage offered by these approaches in the context of pesticides monitoring is that they help us to gain a better understanding of the extent of its uses or contaminations. For instance, the indiscriminate use of pesticides generates serious issues in the environment and human health. Hence, monitoring the use of pesticides is a critical task.

One of the ways to evaluate the levels of pesticides in food that comes to consumer's table is through the indicators of the occurrence of residues. Several countries monitor the occurrence of these residues. For example, in Brazil, this control is executed by the Brazilian Health Regulatory Agency (ANVISA) and in the US by the Department of Agriculture (USDA). The monitoring of residues of pesticides occurs in two ways: (i) direct chemical evaluation of food samples and (ii) prediction through computational simulations of the use of agrochemicals from historical data series [2]. In Brazil, unlike the US, only the first type of monitoring officially occurs.

Simulation-based computing involves several types of paradigms ranging from classical algorithms based on mathematical models to complex multi-objective optimization models supported by Bio-Inspired Algorithms (BIA) and EC [3]. The EC is a family of algorithms for global optimization inspired by biological evolution. Technically, they are a family of population-based trial and error problem solvers with a metaheuristic or stochastic optimization character. EC has as goal to find good enough solutions, without guarantee of the optimal solution, for problems that contain many variables, with multiple targets and present several restrictions. More details about the advantages and disadvantages of using EC over other approaches can found at [3].

Nowadays, the simulations can be performed by scientific workflows. A scientific workflow is an abstraction capable of representing an *in silico* scientific experiment, where the researcher defines the sequence of the programs that will be executed, their parameters and their data dependencies [4].

Scientific workflows are composed and enacted by Scientific Workflow Management System (SWfMS). Although the SWfMS offer a framework for modeling and executing workflows in centralized or distributed environments, they still lack broad support for adoption of EC and BIA.

Supporting the development of scientific workflows that require EC and BIA for the solution of simulation that includes multi-objective optimization is an open research question. In this work, we propose a novel framework named VisPyGMO. The main goal behind VisPyGMO is to (re)use the general paradigm of parallel EC proposed by Izzo *et al.* [5] and to incorporate generic EC and BIA algorithms into the traditional SWfMS.

VisPyGMO was conceived to be easily used by researchers with little knowledge of computational intelligence or optimization; allowing them to take advantage of EC to model predictive workflows that deal with complex optimization problems that involve large volumes of data. The framework was implemented in the Vistrails [6], one of the most common general-use SWfMS in the literature.

2 VisPyGMO

VisPyGMO uses as theoretical foundation the Island and Archipelago paradigm proposed by [5] to encapsulate parallel EC algorithms. The algorithms we have used are Particle Swarm Optimization (PSO), Fish School Search (FSS), Firefly Algorithm (FA), Ant Colony (AC), Artificial Bee Colony (ABC) and Differential Evolution (DE) [7]. These algorithms can be applied to a large family of optimization problems. Besides, the algorithms we used have native support for parallel computing and were fully compatible with PyGMO libraries (http://esa.github.io/pygmo/). We (re)use these libraries due to the following technical reasons. (i) offer more than 20 types of general-purpose optimization algorithms distributed in three categories (heuristic, metaheuristic, and local optimization); (ii) have native support for parallelism via MPI [8]; (iii) the EC modules are easy to be used in the evolutionary workflows, having few parameters to adjust and; (iv) the algorithms are open-source being underpinned by an active community of developers.

The current version of VisPyGMO is being deployed in VisTrails as a package which is composed of several EC modules, each representing an evolutionary algorithm (*e.g.*, PSO, SA, ABC and DE). All EC packages and modules have been implemented in Python and the algorithms in C ++ and Python.

2.1 Proofs of Concept

Despite being one of the world's largest consumer of pesticides, Brazil still does not have trusted open datasets that contain information about the occurrence of pesticide residues in food. Thus, to evaluate VisPyGMO, avoid analytic bias and, to generate predictions based on evolutionary workflows, we used datasets obtained from the USDA[1]. The raw data were scattered over long series of more than 20 years of analysis of the occurrence of pesticide residues (from 1992 to 2015), with a population of over 40 million of analyzes and 670 classes of pesticides in approximately 80 MB of structured data. The datasets contain data on food type, pesticide name, sample collection location, detected quantities, detected concentration, timestamp information, among others.

The proofs of concept aimed to assess the functionality of the VisPyGMO; one of the first steps when choosing an appropriate research method is to clarify the research question [9]. Our research question was: "What are the most common pesticides used in the US and what are their future consumption trends in the next five years?".

The preliminary experiments we have executed can be summarized as follows. They were divided into two parts. In the first, we developed a scientific workflow (not discussed in the abstract) that pro-

[1] https://www.ams.usda.gov/datasets/pdp/pdpdata

duces fault-free series of data and generates all the islands, grouped into agronomic classes of pesticides and food types. After that, we ranked the top ten pesticides (islands) with higher residues in food samples. We detected that malathion was the most common pesticide used in the last twenty years in the USA, its island was composed of 6.954 samples.

The second part is characterized by another workflow (Figure 1). It was developed in VisTrails using VisPyGMO´s modules. We used DE modules to predict the consumption trends of each class of pesticide. We decide to use DE because it is an evolutionary algorithm used in global optimization problems and it is used to find approximate solutions. Finally, to visualize the tendencies, an ARIMA graphic projection was added to the workflow. The experiments were configured with 95% of tolerance in the presentation of the projections.

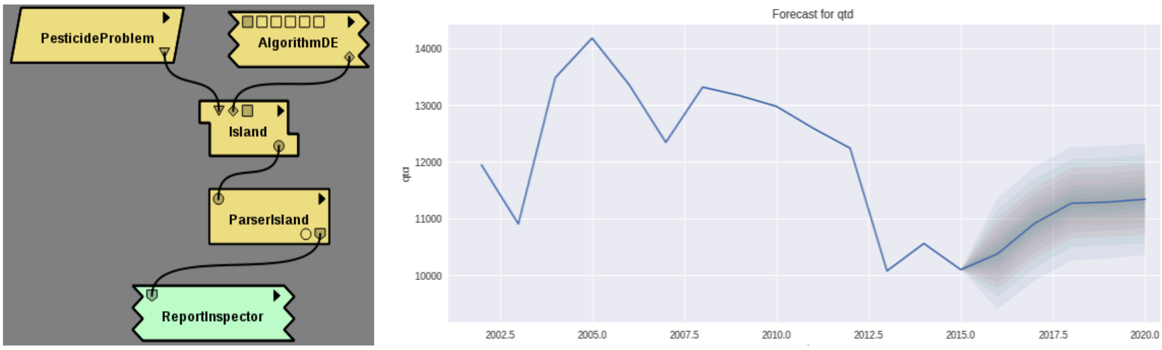


Figure 1: Fragment of an EC workflow in VisTrails using VisPyGMO modules (left). An example of the visualization of usage trends of Malathion in the next five years (right).

3 Conclusion

Investigations that require simulations and predictions based on large datasets are gaining momentum in several areas of science, including agricultural sciences [10]. Our contribution allows researchers the use of evolutionary algorithms in SWfMS in a simplified way. As proof of concept, we developed new reusable EC modules based on the Island/Archipelago paradigm to be used in SWfMS. Also, our initial experiments exploited real datasets to predict future uses of different classes of pesticides in agriculture. As future work is intended to include new optimization algorithms in VisTrails, for example, genetic algorithms, Monte Carlo searches in parallel clusters and cloud environments.

Acknowledgments

This work was supported in part by MEC/FNDE. The authors thank PET-SI/UFRRJ, PPGMMC/UFRRJ, Red CYTED - BigDSSAgro, and to Professor Lluis M. Plá.

References

[1] Stephan Olariu and Albert Y. Zomaya (Eds.), *Handbook of Bioinspired Algorithms and Applications*. Chapman & Hall/CRC - Computer and Information Science Series, 2005.

[2] Xiu-Hong Zhang et al., The Prediction of Pesticides Usage Based on PSO Algorithm. *Int. Conf. Computational Problem-Solving* (*ICCP2012*) pages. 508-511, China. October 19-21, 2012.

[3] Thomas Back, David B. Fogel and Zbigniew Michalewicz, (Eds.) *Handbook of Evolutionary Computation*, Institute of Physics Publishing and Oxford University Press, 1997.

[4] Ewa Deelman et al., Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems,* 25:5, pages. 528–540, 2009.

[5] Dario Izzo, Marek Ruciński and Francisco Biscani, *The Generalized Island Model.* In: Parallel Architectures and Bioinspired Algorithms. Springer Berlin Heidelberg. pages. 151–169, 2012.

[6] Steve Callahan; et al., *VisTrails: visualization meets data management*. In: Proceedings of the International Conference on Management of Data (*SIGMOD2006*), pages. 745-747, USA, 2006.

[7] James Kennedy and Russel Eberhart. *Particle Swarm Optimization*. In Proceedings of the of IEEE International Conference on Neural Networks, Perth, Australia, pages. 1942–1948, 1995.

[8] Marc Snir et al., *MPI: The Complete Reference*. MIT Press, 335 pages. 1996.

[9] Steve Easterbrook et al., *Selecting empirical methods for software engineering research*. In: Guide to advanced empirical software engineering, Springer, pages. 285–311. 2008.

[10] Sergio Manuel Serra da Cruz, Maria L. M. Campos and Marta Mattoso. *Towards a Taxonomy of Provenance in Scientific Workflow Management Systems*. In: SERVICES I, pages. 259-266, USA, 2009.

Spatial variability inside a greenhouse can be modeled with machine learning

Vinícius André Velozo Lopes[1], Felipe Ferreira Bocca[1], Luiz Henrique Antunes Rodrigues[1]

[1] University of Campinas – School of Agricultural Engineering
Av. Cândido Rondon, 501 - Barão Geraldo 13083-875 - Campinas/SP - Brazil
vinicius.lopes@feagri.unicamp.br
felipe.bocca@feagri.unicamp.br
lique@g.unicamp.br

Abstract

The aim of this study was to model the temperature and relative humidity on the 45 sensory points inside the greenhouse, using boosted regression trees. Our results show that internal variability of conditions can be modeled as a function of greenhouses' characteristics and external conditions. Even in the worst scenarios, most models could capture the internal variability, indicating that this approach can be used in evaluation of greenhouse's heterogeneity in different weather conditions and structure configurations. Differences in performance seem to be related to changes in variability across time.

1 Introduction

The demand for food is crescent in a scenario where the population is growing and the resources are increasingly scarce. Protected cultivation associated with technologies that prioritize the usage of resources according to the plant's necessities can be a viable alternative to increase crop yield in a sustainable way. This growth in yield provided by the use of greenhouses is mostly related to the protection from extreme conditions. The use of sensors in greenhouses can also be a factor for gain in production, since the highly monitored environment allows for a more effective control.

Bojacá et al. [1] analyzed the temperature gradient inside a greenhouse and its influence on the yield of the cultivar. They also determined the reliability of the use of geostatistical methods to be applied as a tool for estimating this gradient. Those authors found that the horizontal temperature gradient inside a greenhouse can be as high as the one between the inside and the outside. Climate gradients can cause differences in yield and plant's characteristics, as well as promote an environment that favors the development of diseases [2].

Despite that, often studies regarding greenhouses' environment consider their microclimate to be uniform [3]. The data is often obtained using one or few sensory points in the geometrical center of the greenhouse. Furthermore, works that try to model the interior of these environments usually predict the average of environmental factors such as relative humidity (RH) and temperature [4-5].

Considering that the internal conditions are a function of external environment and of the configuration of the greenhouse, one approach could be use this information to model internal heterogeneity. In this study, we modelled temperature and relative humidity in different locations inside a greenhouse given different external conditions and structure configurations. Using boosted regression trees (BRT), we were able to model the internal conditions across the greenhouse with a maximum mean absolute error (MAE) of 0.51 °C for temperature and 3.33% for the RH.

2 Materials and methods

2.1 Experimental structure

Data was collected between June of 2014 and February of 2015, at a greenhouse installed in the experimental field of the School of Agricultural Engineering of the University of Campinas, Campinas/SP (22° 49' 06" S, 47° 03' 40" W and 635 m above sea level) [6]. The greenhouse had a floor dimension of 117 m², width of 6.4 m, length of 18.3 m and height of 4 m.

A wireless sensor network was deployed with 45 sensors inside the greenhouse, acquiring temperature and RH data. Two weather stations near the experimental field were used to supply the database with the amount of rain, temperatures, RH, wind velocity and direction and radiation. Structural elements, such as the use of plastic cover, zenithal window opening and exhausters were analyzed for five configurations.

2.2 Data analysis

Models were developed using the gbm package from R software. Different models for temperature and RH were created for 5 scenarios using the external conditions and greenhouse configurations (Table 1) as inputs.

Greenhouse Configuration	**Structural element**						**Data acquisition period**
	Lateral cover	**Zenithal window**	**Superior Exhauster**	**Inferior Exhauster**	**Thermo - reflective screen**	**Porous pad**	
1	Plastic	Closed	Off	Off	None	Off	06/15/2014 - 06/24/2014
2	Plastic	Closed	On	On	None	Off	06/26/2014 - 07/05/2014
3	Plastic	Closed	On	On	Installed	On	08/31/2014 - 09/09/2014
4	Insect screen	Closed	On	Off	None	Off	10/14/2014 - 10/23/2014
5	None (Opened)	Opened	Off	Off	None	Off	11/27/2014 - 12/06/2014

Table 1: Structural elements and data acquisition period for the 5 analyzed greenhouse's configurations

The data analyzed presented a temporal structure that needs to be specially handled. We sorted the data in a crescent time order, splitting the data frame in two samples, train and test, not randomly divided. The oldest data was used for training and the newest, for testing. Cross-validation for hyper parameter tuning was performed only on the train sample, using for each of the recursive iteration a bigger amount of data. All the iterations receive the same splitting method already mentioned.

3 Results and discussion

The best models for temperature were obtained in scenarios 4 and 5, with a MAE of 0.24 and 0.36 °C, respectively. The worst models for temperature were obtained in scenarios 1, 2 and 3 and achieved a MAE of 0.46, 0.47 and 0.51 °C, respectively. The best models for RH were obtained in scenarios 2, 3 and 5, with a MAE of 1.75, 1.20 and 1.33 %, respectively. The worst models for RH were obtained in scenarios 1 and 4 and achieved a MAE of 3.33 and 3.30 %, respectively.

Most models were able to model the greenhouses' spatial variability across time. The variability of internal conditions can be viewed in Figure 1, in which the actual and predicted values of temperature and RH are presented for the 45 points. When conditions are homogeneous all curves are near each other in the timestamp and a small vertical spread is observed. The large vertical spread is observed when high variability occurs.

In scenarios 4 and 5, which provided the best temperature models, there was a small variability of temperature across the entire period. This did not happen for scenarios 1, 2 and 3, in which variability changed across the test period. The worst RH models, obtained with data from scenarios 1 and 4, seem to have different sources of error. In scenario 1, high heterogeneity was observed during the day and low variability was observed during the night. The model was not able to capture the variability during the day. In scenario 4, the model failed to predict the RH saturation during the nights, underestimating RH.

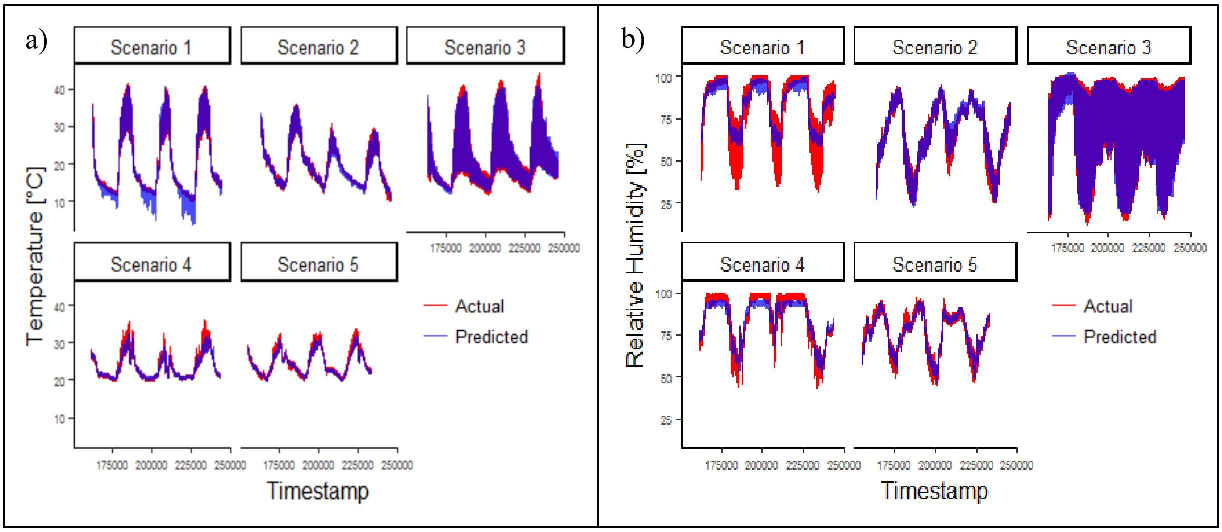


Figure 1: Graphics of greenhouse's internal conditions, actual and predicted, for the 5 considered configurations. a) Temperature. b) RH.

Our results show that internal variability of conditions can be modeled as function of greenhouses' characteristics and external conditions. Even in the worst scenarios, most models were able to capture the internal variability, indicating that this approach can be used in evaluation of greenhouse's heterogeneity in different weather conditions and structure configurations. Differences in performance seem to be related to changes in variability across time.

Acknowledgments

The authors are grateful to Fapesp for their support with the data (Process # 2013/11953-9).

References

[1] Carlos Ricardo Bojacá, Rodrigo Gil, and Alexander Cooman. Use of geostatistical and crop growth modelling to assess the variability of greenhouse tomato yield caused by spatial temperature variations. *Computers and Electronics in Agriculture*, 65 : 219–227, 2009.

[2] Konstantinos P. Ferentinos, Nikolaos Katsoulas, Antonis Tzounis, Thomas Bartzanas, and Constantinos Kittas. Wireless sensor networks for greenhouse climate and plant condition assessment. *Biosystems Engineering*, 153 : 70–81, 2017.

[3] Constantino Kittas and Thomas Bartzanas, "Greenhouse microclimate and dehumidification effectiveness under different ventilator configurations," *Building and Environment.*, 42 : 3774–3784, 2007.

[4] Bram Vanthoor, Cecilia Stanghellini, Eldert Jan Van Henten, and Pieter de Visser. A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates. *Biosystems Engineering*, 110 : 363–377, 2011.

[5] Huihui Yu, Yingyi Chen, Shahbaz Gul Hassan, and Daoliang Li. Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. *Computers and Eletronics in Agriculture*, 122 : 94–102, 2016.

[6] Thais Queiroz Zorzetto. Mapeamento meteorológico de diferentes graus tecnológicos em casas de vegetação. *Doctorate thesis*, pages 36-38, Campinas (SP), BRAZIL, August 28 2015.

Neglecting autocorrelation in development underestimates the error of sugarcane yield models

Matheus A. Ferraciolli[1], Felipe F. Bocca[1], Luiz Henrique A. Rodrigues[1]

[1] University of Campinas – School of Agricultural Engineering
Av. Cândido Rondon, 501 – Barão Geraldo 13083-875 – Campinas/SP – Brazil
matheus.ferraciolli@feagri.unicamp.br
felipe.bocca@feagri.unicamp.br
lique@g.unicamp.br

Abstract

When applying machine learning techniques to agricultural phenomena, data cannot be assumed to be independent, given the potential spatial autocorrelation between samples. This has an impact when creating independent datasets for model assessment. To solve this, samples can be grouped in spatial blocks that are randomly used for splits. In this work, we evaluated the usual and the blocking approaches in the development of sugarcane yield models. The two work ows were evaluated for hyperparameter adjustment, model selection and evaluation. The conventional approach severely underestimated the generalization error and lead to more complex models, potentially over tted to the training data.

1 Introduction

In sugarcane production, the availability of harvesting, meteorological and management data makes it possible to apply machine learning techniques to estimate yield. When using such models, the spatial autocorrelation between samples can be detrimental to several steps during model development, in special, model assessment. Spatial autocorrelation (SA) measures how much the correlation of values of a variable is attributed to its geographic location [1]. In agriculture, adjacent locations have higher probability of presenting similar characteristics, due to soil type and composition, meteorological conditions and crop management being the same for large areas.

The presence of SA in data used for modeling crop yield violates the premise of data independence. This violation leads to similar observations being used for training and testing data, which may result in an optimistic prediction error in the nal model [2]. There are di erent ways in which standard techniques for modeling and organizing data may be modi ed to account for the presence of autocorrelation. A general strategy to increase independence is to split the data into blocks, based on spatial location [3].

In a crop prediction study [2], the data sets for the harvest of two di erent elds were segmented in clusters with approximately the same amount of observations. The spatial cross validation approach used in this work assigned the clusters to the folds, instead of individual observations. Comparing the errors of the spatial and non-spatial setups, the latter led to a substantial underestimation of the prediction error.

In this work, we compared the usual protocol for dividing datasets with the spatial blocks protocol in the development of sugarcane yield models. We used data from three sugarcane mills, with varied areas and eld con gurations, studying the impact of spatial autocorrelation in the separation of folds for training and testing data in hyperparameter adjusting of the model, adapting a previously presented technique [2].

2 Materials and Methods

Our dataset consisted of production data from three mill areas from Odebrecht Agroindustrial, in Brazil: UCR – Usina Costa Rica (Costa Rica – MS), URC – Usina Rio Claro (Caçu - GO) and USL – Usina Santa Luzia (Nova Alvorada do Sul – MS), through the period from 2012 to 2015. The number of observations available for both URC and USL was close to 5000 each, while close to 3000 observations were available to UCR. The dataset included data regarding harvesting, inputs, soil properties and classification, to which we added meteorological information from the area, such as rain, temperature and solar radiation. Dataset construction followed the procedures presented by [4].

The methods for generating models for both approaches were similar, differing only on how we separated data in the steps of training and testing, while adjusting model parameters. We started defining three validation sets: for each mill, we selected groups of fields that were, at least, 3 kilometers apart from nearby fields, in an attempt to reduce spatial correlation between them. In each mill, the validation set consisted of, approximately, 20% of the number of fields used in modeling.

In the standard (STN) approach, we randomly divided observations for training and testing sets. We also randomly assigned each field to different folds for cross validation. In the spatial approach, we used a k-means algorithm to cluster the data and then randomly assigned clusters of fields to training and testing. For the spatialized cross validation (SCV) approach, clusters, rather than individual observations, were assigned for the different folds.

Models were generated by using regression techniques available in R libraries. In both approaches, we utilized Boosted Regression Tree (BRT, from *gbm* package [5]), Support Vector Regression (SVR, from *e1071* package[6]) and Random Forest (RF, from *randomForest* package [7]) to build the models. In order to adjust each technique's hyperparameters we used random search [8]. The configuration with best performance was used to train a model in the training set and further evaluated in the test and validation set.

3 Results and discussion

A summary of the results is displayed on Table 1. Overall, the SCV and STN workflows resulted in different models in each mill for all techniques, except for SVR in USL. Error estimates in the training and testing sets with SCV were higher than STN and closer to the error in the validation set. Both workflows resulted in models with similar errors in the validation set. The performance in the validation set was similar to both approaches, but the STN workflows severely underestimated the error. Also, during model tuning, the STN workflow favored models with increased complexity, like larger number of trees for BRT and RF. The best model for the UCR mill was the RF, while SVR was the best model for both URC and USL. This is similar to previous results for sugarcane yield modeling [4], in which RF and SVR resulted in better models.

When comparing the outcome of both workflows, the resulting error rates behaved according to the expected. The naive approach of random sampling severely underestimated the generalization error, which was best tracked with the SCV workflow. Surprisingly, tracking the error correctly did not enable the SCV work-flow to produce a better model for each mill, despite the technique. Considering the gap between training and validation error in the STN workflow, the models seems overfitted to the training set and this was not correctly tracked by the error in the test set.

Mill	WF	BRT			SVR			RF		
		Train	Test	Val	Train	Test	Val	Train	Test	Val
UCR	SCV	15.71	13.46	17.88	16.5	12.92	18.32	15.41	14.05	17.5
	STN	10.58	9.34	19.94	7.21	9.08	19.68	10.15	8.99	18.05
URC	SCV	12.57	12.84	16.29	12.74	13.35	15.65	12.07	12.65	17.44
	STN	8.16	7.63	16.12	6.54	8.71	16.16	8.28	7.67	17.37
USL	SCV	10.02	14.00	16.10	9.37	13.16	15.08	9.74	13.44	16.69
	STN	6.89	6.35	16.51	4.91	7.24	15.08	7.58	7.03	16.77

Table 1. Mean absolute errors (t.ha^{-1}) in training, testing and validation for models generated by each of the three techniques (BRT, boosted regression trees; SVR, support vector machines; RF, random forest) for the two work ows (WF) being standard (STN) spatial cross validated (SCV). Mean absolute error [t ha^{-1}] is presented for the training, test and validation (Val) sets.

The best model performance for the URC and USL mills were similar (close to 15 t ha^{-1}), while the best model for the UCR mill was higher (17.5 t ha^{-1}). This could be caused for the smaller dataset available for the UCR mill. The gap between test and validation error in the SCV approach could still be related to di erences in the feature spaces in the sets. When the data is divided following a spatial structure, it is possible that certain conditions were not present in both datasets [3].

Our analysis indicated that addressing spatial autocorrelation between observations when developing yield models resulted in a better estimation of the error. Using STN might underestimate the validation error, degrading the models' practical application.

References

[1] D. A. Gri th. Spatial Autocorrelation and Spatial Filtering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

[2] G. Russ.; A. Brenning. Data mining in precision agriculture: management of spatial information. In: Computational intelligence for knowledge-based systems design. [s.l.] Springer, 2010. p. 350–359.

[3] D. R. Roberts; V. Bahn.; S. Ciuti.; M. S. Boyce; J. Elith.; G. Guillera-Arroita; S. Hauenstein; J. J. Lahoz-Monfort; B. Schröder; W. Thuiller; D. I. Warton; B. A. Wintle; F. Hartig; C. F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, n. December 2016, p. 1–17, mar. 2017.

[4] F. F. Bocca; L. H. A. Rodrigues. The e ect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Computers and Electronics in Agriculture, v. 128, p. 67–76, 2016.

[5] Greg Ridgeway with contributions from others (2017). gbm: Generalized Boosted Regression Models. R package version 2.1.3.

[6] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel and Friedrich Leisch (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8.

[7] A. Liaw and M. Wiener (2002). Classi cation and Regression by randomForest. R News 2(3), 18--22.

[8] J. Bergstra; Y. Bengio. Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research, v. 13, n. feb, p. 281–305, 2012.

A multiobjective model to determine the sustainability level of livestock production in the Huascaran National Park

Jesús E. Espinola[1], Henry A. Garrido[1], Ángel Cobo[2], Fernando Salmón[2], Edwin J. Palomino[3], Esmelin Niquín[1], Maximiliano E. Asís[1]

[1] Santiago Antunez de Mayolo University (UNASAM)
Faculty of Science, Huaraz (Perú)
espinolj@gmail.com, gaha1379@gmail.com, esmelinn@yahoo.es, maxasis@hotmail.com

[2] University of Cantabria
Dpt. Applied Mathematics and Computer Science, School of Industrial Engineering, Avda Los Castros s/n, 39005 Santander (Spain)
angel.cobo@unican.es, fersalmon@wanadoo.es

[3] Santiago Antunez de Mayolo University (UNASAM)
Faculty of Environmental Sciences, Huaraz (Perú)
sebasisadi@gmail.com

Abstract

The present work defines a multiobjective model to address the problem of determining the optimal intensity of livestock activity in a protected natural zone of the Peruvian Andes. The model incorporates both economic and environmental aspects and aims to serve as a tool for a sustainable management of the activities of the primary sector in spaces that require special protection.

1 Introduction

Huascaran National Park (HNP) is a protected area in the Peruvian Andes; with an extension of 340.000 Ha, it was declared a protected natural area in 1975 and World Heritage Site in 1985. In the park there are 42 human settlements (small communities) that carry out livestock activities of grazing. The cattle, especially beef cattle and sheep cattle, have been within the protected area since before it was declared a national park; however, the grazing animals within the park constitutes a conflict in the zone. In this context it is necessary a system capable of monitoring the livestock activity in the park and determine the optimal number of animals according to different environmental and economic criteria. The multiobjective decision-making techniques are therefore very suitable for this purpose.

In order to generate and test an optimization model, a small area in the HNP was selected. In this area there are clearly differentiated two grazing zones with particular characteristics that recommend changing the animals from one zone to another depending on the time of year (dry season or rainy season).

2 Model description

The proposed model aims to determine the number of animals that will be in the pastures in each one of the seasons (dry and rainy), the number of animals that will be sold at the end of each season, and the specific moment in which it will occur the change between the two pasture zones. That is, the model considers 73 decision variables:

- t_c is a numeric variable with values between 0 and 365 that represents the time (day of the year) in which the animals will move from the low to the high zone and in which the sale of some animal can also occur. Time $t_0=1$ corresponds to the beginning of the dry season
- X_{ves}: Number of animals of the variety v, age e and sex s that will be grazing at the beginning of the dry season (t_0). In the model two varieties (bovine and ovine) and 6 age groups (0-6 months,

6-12 months, 12-18 months, 18-24 months, 24-30 months, more than 30 months) are considered.

- $V1_{ves}$: number of animals of the variety *v*, age *e* and sex *s* that will be sold at the end of the dry season (t_c).
- $V2_{ves}$: number of animals of the variety *v*, age *e* and sex *s* that will be sold at the end of the rainy season (t_f=365).

There is a set of model parameters that are necessary to define the objective functions: cattle sales prices; fertility rates; probability of birth of females versus males; daily caloric intake required by each type of animal; livestock equivalence in terms of dietary needs of different age groups; maximum percentage of exploitation of the capacity of a pasture zone.

Sustainable development requires implementing suitable policies integrating several competing objectives on economic, environmental, and social aspects. In this case, the model considers economic objectives such as maximizing the sale value of animals (*Obj1*) or the value of the lifestock at the end of the season (*Obj2*). Another objective is related to the need for a balance of the lifestock (female/male ratio, age distribution); in order to do this, an equilibrium index is defined based on the recommendations of experts on the proportions between males and females and between animals of different ages. The objective is the maximization of this index (*Obj3*). Finally, environmental objectives are considered, trying to minimize the environmental impact of the activity in each of the two study zones (*Obj4* and *Obj5*). Consideration should be given to the vegetal cover of each area and the animal feed requirements. To estimate the amount of pasture available in each area, several field trips have been carried out to generate an inventory of the flora of the areas. Satellite images has also been used to estimate the availabilities of grass from the chromatic analysis of the images.

As in any optimization model, a number of restrictions must also be considered, limiting the values of the different variables: temporary restrictions on the migration or sale of animals, restrictions that prevent over-exploitation of grazing areas or equilibrium requirements in lifesotck.

3 Multiobjetive techniques: practical application of the model

Many of the existing multiobjective decision-making techniques are not effective when the number of objectives to consider is high, even when this number is greater than 3. The multiobjective problems with more than 3 functions have been called in the existing literature as many-objective problems, and are characterized by the exponential increase in the number of non-dominated solutions generated by evolutionary techniques that aim to approximate Pareto front. On the other hand, these problems with many objectives require a greater number of solutions to try to approximate this Pareto front in addition to the difficulties in the visualization of the solutions obtained [3]. In this situation, it is necessary to use strategies that reduce the number of objectives without affecting the general structure of the problem [1]. In the context of this work we are exploring the use of two different strategies:

1. Analysis of correlation coefficients on the space of objectives in a sample of feasible points in order to select the two most conflicting criteria. Later, the use of a multicriteria evolution algorithm based on PSO (Particle Swarm Intelligence) allows to obtain an approximation of the Pareto front [2].
2. AHP (Analytic Hierarchy Process) approach [4] to generate a multi-level hierarchical structure of objectives. A set of pairwise comparisons is used to obtain the weights of importance of the decision criteria, and to define a unique objective function. In order to carry out the relative comparison of the importance of the different economic and environmental objectives, we have used valuation questionnaires for agents involved and with different profiles (environmental experts, farmers, policy makers).

This preliminary work wants to present the formulation of the problem, which is in working process still. So far, an important effort has been made to obtain information, generate environmental audits of study areas and construct a multiobjective decision model that integrates the different factors presented. The developed model is presented in this work and allow the generation of a computational tool that can perform simulations of the impact of livestock activity on the study area, raising the awareness of farmers in the area to promote the sustainability of their production, both economically and environmentally. Those responsible for managing the HNP are also provided with an instrument that can support them in their decision-making.

References

[1] Brockhoff, D.; Zitzler, E. Are All Objetives Necessary? On Dimensionality Reduction in Evolutionary Multiobjetive Optimization. In *Parallel Problem Solving from Nature - PPSN IX*. *Lecture Notes in Computer Science*, vol 4193, pp 533-542, 2006.

[2] Coello, C.A.; Salazar, M. (2002). MOPSO: a proposal for multiple objective particle swarm optimization. CEC'02 Proceedings of the Evolutionary Computation on 2002. Volume 02, pages 1051-1056. ISBN:0-7803-7282-4.

[3] Deb, K.; Saxena, D.K. (2005). On finding Pareto-Optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. KanGAL Report num. 2005011. Kanpur Genetic Algorithms Laboratory (KanGAL). Indian Institute of Technology Kanpur.

[4] Saaty, T.L. The Analytic Hierarchy Process. McGraw-Hill, New York, 1980.

Using a multiobjective DEA model to assess the eco-efficiency of organic blueberry orchards in the CF+DEA approach

Lidia Angulo Meza[1], João Carlos Soares de Mello[2], Alfredo Iriarte[3], Marcela González Araya[4], Ricardo Rebolledo-Leiva[5]

[1,2] Universidade Federal Fluminense
Rua Passo da Patria, 156, São Domingos, Niteroi – RJ, Brazil
lidia_a_meza@pq.cnpq.br, jcsmello@producao.uff.br

[3,4,5] Universidad de Talca
Camino a los Niches km 1, Curicó, 3340000, Chile
mgonzalez@utalca.cl, airiarte@utalca.cl, rrebolledo10@alumnos.utalca.cl

Abstract

In the LCA+DEA approach different DEA models are used to assess the eco-efficiency of units and to obtain a single target for each inefficient unit. A multiobjective DEA model (MORO) provides a set of targets for each inefficient unit, from which the decision maker can choose. This model is used to assess the eco-efficiency of organic blueberry orchards and their set of targets. The results are discussed and compared with those using the BCC model and as well as the targets for the inefficient units. The advantages of the MORO over the single objective DEA models are discussed.

1 Introduction

LCA+DEA approach has been used to determine the eco-efficiency of units taking into account both the operational and environmental aspects of units (services and processes). Besides the eco-efficiency index for all units (identifying best practices), DEA models provide targets for inefficient units to become efficient. A review of applications and approaches for the LCA+DEA can be found in [1, 2]. In this paper, a multiobjective DEA approach is used, a set of models called the MORO models, which provide a set of targets for each inefficient DMUs instead of only one target. A set of targets provide flexibility to the decision makers (the manager or owner) from where to choose according to their operational/managerial needs or possibilities. As the MORO models do not provide an efficiency index for each target, the Index based on vector properties [3] will be used to determine the efficiency of each target. This index can also be used to choose the better target for an inefficient DMU. The same group of organic blueberry orchards, as well as the four-step approach for LCA+DEA from [4], used to implement the LCA+DEA. to illustrate the use of the model and a discussion of results and comparison will be made.

2 Material and Methods

In the four-step approach proposed by [4], the first step is the LCA inventory (i.e. the LCI) and the second step is the environmental characterization (only the carbon footprint estimation of the orchards is considered in the study). This two steps were performed by [5] in a LCA evaluation of five organic blueberry producers (identified by A, B, C, D, E) from central Chile and from three different harvest seasons, 2011/2012, 2012/2013 and 2013/2014. As it was not possible to obtain all data for the 2011/2012 season for producer A, that season for this producer is not considered.

The third step is eco-efficiency assessment using and output oriented DEA model (efficiency indexes, best practices and targets). In this paper, a multiobjective DEA model, called MORO [6, 7] will be used that obtains a set of targets providing this way flexibility to the orchads. As in [4] it will be used as inputs Fertilizers (kg/season) and Energy (MJ/season); and as outputs Production (kg/season) and Carbon Footprint – CF (kg CO_2-eq/season), this last one as an undesirable output (a product of the process to minimize). The MORO model for the four variables in this study is presented in (1), where Production is a variable to maximize and CF is a variable to minimize.

$$
\begin{aligned}
&\text{Max } \phi_0 \\
&\text{Min } \varphi_0 \\
&\text{subject to} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1)\\
&\sum_{j=1}^{n} y_{Pj}\lambda_j = \phi y_{P0};\ \sum_{j=1}^{n} y_{CFj}\lambda_j = \varphi y_{CF0};\ \sum_{j=1}^{n} x_{ij}\lambda_j \le x_{i0},\ i = \text{fertizers, energy};\ \phi_0, \varphi_0, \lambda_j \ge 0
\end{aligned}
$$

In this model ϕ is the factor to increase Production, φ the factor to reduce CF, λ_j is the contribution intensity of benchmark j to the target of the observed orchard-season, x_{ij} is the input i of DMU j, y_{Pj} and y_{CFj} are the Production and CF of DMU j, respectively; 0 refers to the DMU under evaluation. Two additional restrictions are introduced to ensure that the output production does not reduce ($\phi \geq 1$) or the undesirable output, CF, does not increase ($\varphi \leq 1$), this is called the MORO-D. Moreover, as in [4] a variable returns to scale (VRS) will be consider by adding the restriction $\sum \lambda_j = 1$, the model will be called the MORO-D-VRS model. This model is used for each orchard-season. As the MORO models do not provide an efficiency index, the index based on vector properties by [3] is used to determine an efficiency index for each target provided by the MORO-D-VRS model. For this case study, the efficiency index, h, of each target is presented in (3). The fourth step of the approach is the determination of the factors targets, factor that contribute to CF, which is done following the procedure presented in [4] that used the best practiced and targets determine the previous step.

$$h = 1 - \sqrt{(1-\varphi)^2 + (1 - 1/\phi)^2} \qquad (3)$$

3 Results, discussion and conclusions

The MORO-D-VRS model is used to evaluate 14 orchards-season, when an orchard is efficient both ϕ and γ are equal to one, meaning that no changes are required for these variables. For the inefficient orchard-seasons, the model found from 2 to 4 different solutions/targets. Table 1 presents the set of targets for the inefficient orchard-season E 13-14.

Solution	Variations for the variables		h index (%)	Targets levels		Benchmarks
	ϕ	γ		Production (kg/season)	CF (kg CO_2-eq/season)	
1	1,75866	1	56,86	144738	40327	A 13-14, B 12-13
2	1	0,476809	47,68	82300	19228	C 11-12, D 12-13
3	1,28797	0,630843	56,84	106000	25440	A 13-14
4	1,03281	0,484787	48,38	85000	19550	D 12-13

Table 1: Results for the inefficient orchards using the MORO-D-VRS for orchard-season E 13-14.

Table 1 presents four different targets. This table also presents their eco-efficiency index (h index), and their respective levels to be achieve to become efficient. It is worth to point out that all efficient orchard-season deemed efficient in [4] are also efficient using the MORO-D-VRS model. Once the targets for Production and CF are obtained, the targets for the factors that contribute to the CF are determined and presented in Table 2. These new levels are determined using the benchmarks (best practices) and their intensities for each target to achieve the target CF determine by the MORO-D-VRS model.

	Fertilizers (kg/season)	Pesticides (kg/season)	Energy (MJ/season)	Machinery (kg/season)	Materials (kg/season)	Packaging Residues (kg/season)
Current	39119	791	135972	127	444	129869
Solution 1	34804	521	162686	128	2992	181918
Solution 2	5106	313	182637	111	1197	35919
Solution 3	14117	974	144719	159	3853	160574
Solution 4	3442	354	201699	111	1273	39615

Table 2: Targets for the factors that contribute to CF for the orchard-season E 13-14

From this set of target, the orchard-season E 13-14 can choose the target in different ways: the most efficient, the one that prioritizes production, the one that prioritizes CF, the one that is a compromise between increasing Production and reducing CF, the one that represents a feasible variation in the production and the resources used in the process, or the one that has only one benchmark. The most efficient target, solution 1, maintains the CF but needs to increase its production in nearly 76% more. There are two benchmarks, A 13-14 and B 12-13, so the targets for the factors in table 2 follow the practices of these benchmarks and their intensities. The one that prioritizes Production is coincidently the same as before. The one that prioritizes CF is solution 2, which has to reduce its CF in nearly in half, that has two benchmarks, C 11-12 and D 12-13, and its respective targets in table 2 follow the practices from these benchmarks and their intensities. The one that is a compromise between variations in production and CF are solution 3 and 4. The one that represents the feasible variations in the variables depend on the characteristics of the orchards, targets levels have to be verified and determine if they are feasible from a managerial/operational point of view. Finally, the one that has only one benchmarks has a relevance in many instances where it is difficult to embody practices from more than one benchmark, which one is to be followed? In these cases, it is preferable to choose a solution that has only benchmark, like solutions 3 or 4, in this way guidelines become clearer and change may become easier to implement.

References

[1] Ian Vázquez-Rowe and Diego Iribarren. Review of Life-Cycle Approaches Coupled with Data Envelopment Analysis: Launching the CFP + DEA Method for Energy Policy Making. *The Scientific World Journal,* 2015: 10., 2015.

[2] Mario Martín-Gamboa, Diego Iribarren, Diego García-Gusano, and Javier Dufour. A review of life-cycle approaches coupled with data envelopment analysis within multi-criteria decision analysis for sustainability assessment of energy systems. *Journal of Cleaner Production,* 150: 164-174, 2017.

[3] Silvio Figueiredo Gomes Junior, João Carlos Correia Baptista Soares de Mello, and Lidia Angulo-Meza. DEA nonradial efficiency based on vector properties. *International Transactions in Operational Research,* 20: 341-364, 2013.

[4] Ricardo Rebolledo-Leiva, Lidia Angulo-Meza, Alfredo Iriarte, and Marcela C. González-Araya. Joint carbon footprint assessment and data envelopment analysis for the reduction of greenhouse gas emissions in agriculture production. *Science of The Total Environment,* 593–594: 36-46, 2017.

[5] Hanna Cordes, Alfredo Iriarte, and Pablo Villalobos. Evaluating the carbon footprint of Chilean organic blueberry production. *International Journal of Life Cycle Assessment,* 21: 281-292, 2016.

[6] Marcos Pereira Estellita Lins, Lidia Angulo-Meza, and Angela Cristina Moreira da Silva. A multi-objective approach to determine alternative targets in data envelopment analysis. *Journal of the Operational Research Society,* 55: 1090–1101, 2004.

[7] João Quariguasi Frota Neto and Lidia Angulo-Meza. Alternative targets for data envelopment analysis through multi-objective linear programming: Rio de Janeiro Odontological Public Health System Case Study. *Journal of the Operational Research Society,* 58: 865–873, 2007.

Using CF+DEA method for assessing eco-efficiency of Chilean vineyards

Ricardo Rebolledo-Leiva[1], Carlos Rodríguez-Lucero[2], Melany Campos-Rojas[2], Eduardo Pacheco-Rojas[2], Marcela C. Gonzalez-Araya[3], Alfredo Iriarte[3], Lidia Angulo-Meza[4]

[1] Centro de Extensionismo Tecnológico de Logística, Universidad de Talca
Camino a los Niches km 1, Curicó, 3340000, Chile
ricardo.rebolledo@cetlog.cl

[2] Escuela de Ingeniería Civil Industrial, Facultad de Ingeniería, Universidad de Talca
Camino a los Niches km 1, Curicó, 3340000, Chile
carodriguez12@alumnos.utalca.cl, mcampos12@alumnos.utalca.cl, epacheco12@alumnos.utalca.cl

[3]Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad de Talca
Camino a los Niches km 1, Curicó, 3340000, Chile
mgonzalez@utalca.cl, airiarte@utalca.cl

[4] Universidade Federal Fluminense
Rua Passo da Patria, 156, São Domingos, Niteroi – RJ, Brazil
lidia_a_meza@pq.cnpq.br

Abstract

The agricultural activity has impact on environmental sustainability. The eco-efficiency plays a fundamental role. This concept is defined as the delivery of goods, using fewer resources and decreasing environmental impacts. In this research, we assess the eco-efficiency of Chilean vineyards with the joint use of Carbon Footprint (CF) and Data Envelopment Analysis (DEA). After determining the GHG emissions, basic DEA models were used to determine the eco-efficiency (CCR and BCC models). The results show few efficient vineyards. For inefficient vineyards, DEA models set targets for increasing grape production and reducing agrochemicals consumption.

1 Introduction

The joint implementation of LCA and DEA can assess the operational and environmental performance of multiple units. This approach has the advantage of detecting and changing technical inefficiencies that are the source of unnecessary environmental impact. LCA propose indicators to evaluate different impacts to environment. The CF is one of these indicators that seeks to assess the GHG emissions that contribute to Climate Change. Recently, the CF + DEA approach was proposed by [6] as a combination of CF and DEA to benchmark operational and environmental performance in terms of GHG emissions in energy entities. In this research, the CF + DEA approach proposed by [5] is applied for evaluating the eco-efficiency of nine Chilean vineyards. This evaluation follows the eco-efficiency definition, which is to deliver goods using fewer resources and reducing the environmental impact.

2 Material and Methods

For the eco-efficiency assessment, we used the four-step method proposed by [5]. In the first step, the Life Cycle Inventory (LCI) is estimated. In this step, it is necessary to collect the inputs and outputs data relevant for the system. Data of 2015/2016 season from nine vineyards located in the Chilean Central Valley were collected. The second step calculates the carbon footprint (CF) of the nine vineyards. In this step, 1 kg of harvested grape is considered as functional unit. Besides, the system boundary is set from cradle-to-farm gate. The method used to evaluate the CF for each vineyard follows the ISO 14040 general framework [4] and the CF is calculated according to PAS 2050 standard [1] with its specification for horticulture PAS 2050-1 [2]. The assessment of CF was modeled in Ccalc2 v1.43. The third step corresponds to eco-efficiency assessment using output oriented DEA models. In this evaluation, CCR and BCC models are used. The efficiency index, best practices and targets are identified. The inputs considered for each vineyard are: Fertilizers (kg/season), Pesticides (kg/season). The outputs are: Production (kg/season), CF (kg $CO_{2\text{-}eq}$/season). The CF is an undesirable output, so we seek to minimize. We propose to deal with undesirable outputs using the multiplicative inverse transformation proposed by [3]. In this way, we are able to identify efficient units that maximize production with low CF emissions. The DEA models were implemented in IBM ILOG CPLEX Optimization Studio 12.6 on an Intel® CoreTM i5-337U processor, operating at 1.80 GHz. Finally, the last step establishes a method to determine the factor targets in order to achieve CF reduction. This step proposes to replicate benchmark practices (the best practices of real farmers), for each inefficient vineyard.

3 Results, discussion and conclusions

For each vineyard, the data were obtained in several face-to-face interviews. The non-productive stages of the crop (e.g., planting and growing) and pruning residue treatments (burning, mulching, etc.) are excluded from this evaluation. The obtained results in the second step are presented in Figure 1. In this figure, it is possible to note that, in average, 50% of CF comes from fertilizers, while 46% corresponds to pesticides. Only 4% is contributed by energy.

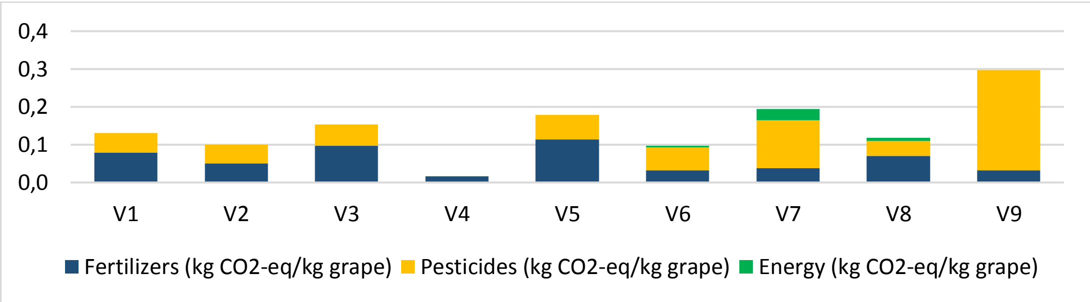


Figure 1. Contribution of the factors in the vineyards' carbon footprint

The efficiency index obtained in the third step using the CCR and BCC models is presented in Table 1. The BCC model identifies three efficient vineyards (V4, V7, V8), while CCR model identifies only one efficient vineyard (V4).

Vineyards	V1	V2	V3	V4	V5	V6	V7	V8	V9
Efficiency CCR	0.2	0.3	0.2	1	0.1	0.5	0.4	0.2	0.5
Efficiency BCC	0.8	0.5	0.6	1	0.6	0.9	1	1	0.7

Table 1. Efficiency indexes of CCR and BBC models

Since BCC model considers scale and operation differences, we use only this model for estimating vineyard targets. Production and CF targets for inefficient vineyards are shown in Table 2.

Inefficient Vineyards	Production (kg)	Production Target (kg)	CF (kg $CO_{2\text{-}eq}$)	CF Target (kg $CO_{2\text{-}eq}$)
V1	147049	180968	19195	15597
V2	23436	50820	2348	716
V3	72650	120954	11143	1186
V5	131847	217207	23504	12290
V6	63063	70422	23504	769
V9	44547	60762	6124	733

Table 2. Production and CF targets

The factors targets for CF reduction are shown in Table 3. According to this table, an average reduction about 59% of fertilizers is proposed, while an average decrease around 81% of pesticides is estimated.

Inefficient Vineyards	Fertilizers (kg)	Target of Fertilizers (kg)	Pesticides (kg)	Target of Pesticides (kg)
V1	6000	5312	797	441
V2	610	331	111	4
V3	3644	472	415	20
V5	7832	3769	821	403
V6	1068	338	393	7
V9	731	59	890	9

Table 3. Factor target for CF reduction

As expected, the CCR efficiency scores were lower than the BCC's, allowing to establish a ranking for nine vineyards. However, we do not suggest the CCR model application, because this model assumes constant returns to scale, which means that an increase of inputs will produce a proportional increase in the outputs. This proportional increase is constant independently of the size or scale in which the vineyards operates.

Finally, we expect that inefficient vineyards follow the operational and managerial guidelines of their related benchmarks. When an inefficient vineyard has more than one benchmark, it is necessary to identify which ones have greater intensities, because that means these benchmarks have similar characteristics than the inefficient vineyard.

References

[1] BSI, 2011. PAS 2050:2011. Specification for the Assessment of Life Cycle Greenhouse Gas Emissions of Goods and Services. British Standards Institution, London.

[2] BSI, 2012. PAS 2050-1:2012. Assessment of Life Cycle Greenhouse Gas Emissions from Horticultural Products. British Standards Institution, London.

[3] Golany, B., Roll, Y., 1989. An application procedure for DEA. Omega 17 (3), 237–250.

[4] ISO, 2006. ISO 14040:2006. Environmental Management—Life Cycle Assessment—Principles and Framework. International Organization for Standardization, Geneva.

[5] Rebolledo-Leiva, R., Angulo-Meza, L., Iriarte, A., González-Araya, M.. Joint carbon footprint assessment and data envelopment analysis for the reduction of greenhouse gas emissions in agriculture production. Science of The Total Environment, 593–594: 36-46, 2017.

[6] Vázquez-Rowe, I., Iribarren, D. Review of Life-Cycle Approaches Coupled with Data Envelopment Analysis: Launching the CFP + DEA Method for Energy Policy Making. The Scientific World Journal, 2015: 10, 2015.

Review of Data mining applications in forestry sector

Broz, Diego[1]; Olivera[2], Alejandro; Viana Céspedes, Víctor[2]; Rossit, Daniel Alejandro[3]

[1] UNaM CONICET Facultad de Ciencias Forestales
Bertoni 124, Eldorado (N3382GDD), Misiones, Argentina.
diego.broz@facfor.unam.edu.ar

[2] Universidad de la República
Ruta 5, km 386,5, 45000 Tacuarembó, Uruguay.
alejandro.olivera@cut.edu.uy, vviana@cci.edu.uy

[3] Departamento de Ingeniería, Universidad Nacional del Sur, CONICET
Av. Alem 1253, Bahía Blanca (B8000CPB), Argentina.
daniel.rossit@uns.edu.ar

Abstract

Modern technology makes possible to collect large amount of data that can be processed and transformed in valuable information for several human activities. Forest industry particularly can take advantage of such technology because of modern forest harvesters are equipped with a system for data collection and communication called StanForD. Data mining allows users to process large databases to determine trends and patterns. In this extended abstract we present a brief revision of the literature dedicated to the issue and, also, we indicate synthetically future research directions that could be useful for forest operations management. Some DM techniques are artificial neural network and decision tree and they are used to perform association, classification and clustering. Nonetheless, data mining techniques have been successfully applied to several fields, e.g. industry, marketing, sociology, economy, agriculture and environmental sciences.

1 Introduction

The amount of data recorded and stored daily is growing due to the use of technology with automatic data collection capabilities. By processing this data trends and patterns can be determined to use as input in the decision making process in many activities.

Given the amount of data available, to transform the data into information it is necessary to use special techniques that can handle and process the data. Data mining (DM) techniques arise as the solution to the problem. DM is the process of applying Computer Based Information Systems (CBIS) for discovering knowledge from data [1]. DM applications started in the 1960s, its baseline is grounded by disciplines such as machine learning, artificial intelligence, probability and statistics [2].

Knowing the impact variables on the forest harvesting productivity, transport of forest products, plantations establishment, silvicultural practices, among others factors, would make possible to manage such variables more efficiently across the forest industry. In addition, the technology used for harvesting in forest plantations in countries like Uruguay provides a mechanism to automatically record data from forest harvesters during the operation. In this sense, the DM allows to discover patterns and trends useful to predict the behavior of a system and interrelations of interest [3]. Thus, it would be a strategical issue to process that automatically collected data in order to support the decision making process. In this extended abstract we present a brief literature review of data mining implementations on forestry systems, and then, some new possible implementations with impact in forestry management are discussed.

2 Literature review

Different techniques are used for DM, e.g. artificial neural network, regression trees and decision tree (DT), and others [2, 4]. DM has been successfully applied to industrial processes manufacturing, marketing, sociology and others; however, there is little evidence of application to forestry, environmental sciences and agriculture. In [5], authors compared linear and regression tree analyses for forest attribute estimations and their spatial modeling. The results of analysis showed that, statistical models of stand volume, tree density, species richness and reciprocal of Simpson indices using tree regression analysis had higher adjusted compared to linear regression models. Using DM techniques, [6] estimated the risk on forest fire and some of the methods analyzed were multilayer perceptron, radial basis function networks, support vector machines and fuzzy logic. They used historical forest fire records, which contained parameters like geographical conditions of the existing environment, date and time when the fire broke out, meteorological data such as temperature, humidity and wind speed, and the type and tree stocking. In [7], the authors used a DM methodology named instance-based classification for estimating carbon storage in *Araucaria angustifolia* (Bertol.) Kuntze plantation in Brazil. They concluded that the technique outperformed the conventional methods.

In [8], authors examined and analyzed three European projects as guidance to describe current possibilities and future challenges for deployment of Big Data (BD) techniques in the field of agro-environmental research, facilitating decision support at the level of societal challenges. The authors recommended the use of BD to analyse data from various sources, e.g. harvesting, production, and meteorological records. [9] were pioneers on integrating GNSS with forest harvester data to improve forest management. They retrieved data from a GNSS-enabled harvester working in cut-to-length operations in *Eucalyptus spp*. plantations in Uruguay. The dataset obtained comprised over 63,000 cycles of felled and processed stems. With this data, a mixed effects model was fitted to evaluate harvester productivity as a function of stem diameter at breast height, species, shift, slope and operator. To analyze the relevant economic, social and ecological factors of China's forestry resources [10] used the BD theory. Firstly, the authors used the method of data envelopment analysis to investigate the forestry resources efficiency; then they analyzed time series data using the Malmquist total factor productivity index method. Applying Neural network-based models [11] presented a large-scale evaluation of climate effects on the productivity of three temperate tree species in Central Europe. Using this technique they determined which among 13 tested climate variables best predicted the tree species-specific site index. To the best of our knowledge there is no evidence of studies using DM techniques applied to forest harvesting operations or data automatically collected by harvesting machines. As such, there is a potential field of application of DM techniques and compare the results against conventional Regression analysis as performed [9].

3. Conclusion

Various DM techniques have been applied in research for the agro-environmental sciences, including forestry. Prediction of forest fires, the effect of climatic variable on forest productivity, forests structure analysis and carbon storage are some of the case studies published. Techniques comprise mixed models, artificial neural network, association rules, and regression trees. However, there is still a gap regarding the use of DM techniques in forest operations, concretely using the automatic data collection system available in the majority of modern forest harvesters. This data enables to describe internal processes of the system based on actual data. Following these ideas, interesting future research could be oriented to estimates internal parameters that have significant impact on forestry operations planning, such as productivity rate (volume of harvested wood per hour, for example) and operations time (processing and transport times). In a more strategic management perspective, new maps can be developed. Analysing data from past harvesting campaigns, new fit-for-purpose forest yields maps can be built. Also, it is possible to assess and redefine internal forest roads according to the real land and site yield.

References

[1] Vlahos, G. E., Ferratt, T. W., Knoepfle, G. The use of computer-based information systems by German managers to support decision making. *Information & Management*, 41(6), 763-779, 2004.

[2] Jothi, N., Husain, W. Data mining in healthcare–a review. *Procedia Computer Science*, 72, 306-

313, 2015.

[3] Ahlemeyer-Stubbe, A., Coleman, S. *A practical guide to data mining for business and industry*, John Wiley & Sons, 2014.

[4] Liao, S. H., Chu, P. H., Hsiao, P. Y. Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications*, 39(12), 11303-11311, 2012.

[5] Mohammadi, J., Shataee, S., Babanezhad, M. Estimation of forest stand volume, tree density and biodiversity using Landsat ETM+ Data, comparison of linear and regression tree analyses. *Procedia Environmental Sciences*, 7, 299-304, 2011.

[6] Özbayoğlu, A. M., Bozer, R. Estimation of the burned area in forest fires using computational intelligence techniques. *Procedia Computer Science*, 12, 282-287, 2012.

[7] Sanquetta, C. R., Wojciechowski, J., Dalla Corte, A. P., Rodrigues, A. L., Maas, G. C. B. On the use of data mining for estimating carbon storage in the trees. *Carbon balance and management*, 8(1), 1-9, 2013.

[8] Lokers, R., Knapen, R., Janssen, S., van Randen, Y., Jansen, J. Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, 84, 494-504, 2016.

[9] Olivera Farias, A. *Exploring opportunities for the integration of GNSS with forest harvester data to improve forest management*. PhD thesis, University of Canterbury, Nueva Zelanda, 2016.

[10] Li, L., Hao, T., Chi, T. Evaluation on China's forestry resources efficiency based on big data. *Journal of Cleaner Production*, 142, 513-523, 2017.

[11] Hlásny, T., Trombik, J., Bošeľa, M., Merganič, J., Marušák, R., Šebeň, V., Štěpánek, P., Kubišta, J., Trnka, M. Climatic drivers of forest productivity in Central Europe. *Agricultural and Forest Meteorology*, 234, 258-273, 2017.

|Application of data mining to forest operations planning

Rossit, Daniel Alejandro[1]; Olivera [2], Alejandro; Viana Céspedes, Víctor[2]; Broz, Diego[3]

[1] Departamento de Ingeniería, Universidad Nacional del Sur, CONICET
Av. Alem 1253, Bahía Blanca (B8000CPB), Argentina.
daniel.rossit@uns.edu.ar
[2] Universidad de la República
Ruta 5, km 386,5, Tacuarembó (C.P. 45000), Uruguay.
alejandro.olivera@cut.edu.uy, vviana@cci.edu.uy
[3] UNaM CONICET Facultad de Ciencias Forestales
Bertoni 124, Eldorado (N3382GDD), Misiones, Argentina.
diegoricardobroz@gmail.com

Abstract

In Uruguay, mechanized forestry harvesting for industrial purposes is carried out using modern equipment. They are capable of record a wealth of information that can be exploited in the decision making process and improve operations. Some approaches from data mining field, as decision trees, are an alternative to analyze large volumes of data and determine incidence factors. In this work, it was proposed to analyze how different variables of the forest harvest (DBH [1], species, shift and operator) affect the productivity of a forest harvester. Data were collected automatically by a forest harvester working on plantations of *Eucalyptus spp*. in Uruguay. The results show that DBH is the most influential factor in productivity.

1 Introduction

Forest planning is a highly complex decision-making problem involving various factors: ecological, productive and economic systems ([1], [2], [3]). A large extent of this complexity is due to the duration of biological processes involved, such as tree growth, since the length of rotations can reach 25 years [4]. This makes the planning of harvest operations complex and affects the economic performance of companies. Forest harvesting is key factor in commercial forest plantations because of its high impact on production costs, quality and value recovery of forest products (mainly wood) and, also, on their potential environmental impact. In this sense, estimating the productivity (measured in m^3h^{-1}) of these activities is a central issue for planning harvesting operations efficiently. Therefore, a precise estimation of harvester productivity will contribute to improve the supply chain of forest products (from the field to the industry).

Forest harvesting operations in Uruguay use modern machines equipped with automatic data collection technology. This fact makes available a large amount of harvest data that can be processed using data mining techniques for later use in harvest planning and forest management. Olivera et al [5] studied the productivity of harvesting operations in *Eucalyptus spp*. Plantations in Uruguay using data automatically collected by a harvester. With this data, the authors performed a regression analysis to study the effect of five variables on the machine productivity. The variables that significantly affected productivity in order of importance were: Diameter at Breast Height (DBH) of the trees, operator, and work shift (day and night). However, the regression analysis method only allows comparing the dependent variable productivity with a single independent variable at a time, something that limits a more integrative view of the system. In this paper we propose to revisit this problem, using a data mining approach, specifically, classification or decision trees (DT). This methodology will allow a more accurate description of the dependent variable productivity by analyzing its dependence on a set of variables at a time, instead of a single variable. According to Ahlemeyer-Stubbe and Coleman [6], DTs are popular and reliable methods for developing prediction and classification models. For the best of our knowledge, there is no evidence of the application of this technique in forest harvesting planning, although it is a very versatile technique for exploratory data analysis. The objective of this work is to apply this technique using a data set collected automatically by a forest harvester to evaluate the productivity of the operation.

1 DBH: diameter at breast height.

2 Metodologhy

DT methodology is widely disseminated in the field of data mining. It consists of generating a prediction model of a dependent variable as a function of a set of independent variables. The generated model is a tree, in which each branch describes rules in terms of the independent variables that allow to predict categories of the dependent variable with a good level of approximation. This model is based on the exploration of a set of observations. In this paper, we use the classification tool of the SPSS IBM software, and CHAID (Chi-square Automatic Interaction Detector) as analysis procedure.

Our case study comprises a data set of 4805 records of processed trees, obtained from data collected by a forest harvester working on plantations of *Eucalyptus spp*. The machine registers a time stamp when it each tree is fall. We calculated the cycle time of each processed tree determining the difference between two consecutive records as explained in [5]. In addition to the time stamp, the machine also records for each tree: harvested volume (m^3) and DBH. Complementary information was included as variables that can affect productivity: species, shift (day / night) and operator. Productivity was the dependent variable, which was converted into categorical variable, where each category indicates a range of productivity. Next, the decision tree method formulates rules to predict the occurrence of each productivity category. In this work, we propose a gross categorization of productivity to be able to present the methodology. These categories are too broad for a real practical purpose, but to discretize in lower range would imply a larger number of categories, which would turn this work little illustrative and a cumbersome example. The categories adopted are 4 and were named by their upper bound, the first is "<12" considering productivities below 12 m^3h^{-1}, the second: "<= 26" for productivities between 12 and 26 m^3h^{-1}, the third: "<= 40" for productivities between 26 and 40 m^3h^{-1}, and the last one, "> 40" for productivities that exceed 40 m^3h^{-1}.

Once the decision tree model is validated, it can be used to predict the forest harvester productivity in similar situations. This prediction is valuable for planning forest operations and, consequently, for forest product supply chain management.

3 Results and Discussion

Figure 1 illustrates the results obtained by DT models. In the DT model, the dependent variable is Productivity The node 0, shows the observations obtained for each category. Then, the first branch uses the variable "DBH" to classify the observations, generating 8 nodes. From nodes 1 to 8 in each node there is a dominant category of productivity, and, each dominant category has a percentage greater than the percentage at node 0 (greater purity). For example, in the node 1, where DBH is below or equal to 122 mm, the most likely category of productivity is "<12" with almost 80% probabilities. At the next level, nodes 6 and 7 are branched and the variable "Operator" is used to classify the observations. At this level "Operator" values are indicated for each sheet, where the purity of the nodes is improved, as in node 9, which allows to improve the productivity prediction "> 40", from 41.6% to 51.2%. Then, the prediction rules are read from the leaf node to the root node. For example, the rule for node 9 (a leaf node) is: if "Operator" is operator 1, and the DBH is between 204 and 235 mm the productivity will probably be "> 40". In contrast, for node 10, the difference would be "Operator" is operator 2, then the productivity is likely to be "<= 40".

The "Shift" variable did not have a significant effect, so productivity does not significantly depend on the shift or, at least, does to a lesser extent than "DBH" and "Operator". This agrees with [5] and [7], which validates the proposed method, at least, in a gross manner for this brief instance.

The DT method allowed establishing differences in productivities between operators (1 and 2) for DBH classes between 204 and 274 mm for *E, dunnii*. This result allows a more accurate quantification of productivity differences between operators than the regression analysis presented in [5] and [7]. In the mentioned work, the significant difference between operators is established, but the methodology (regression analysis) did not allow to detail for which diameter classes that difference was significant. This level of detail if possible by applying the DT technique. In practice, this level of detail allows to allocate operators to work in different strata of a plantation of this species of *Eucalyptus* according to the expected productivity. In strata with diameters smaller than 204mm, either operator will have similar productivity, while above this diameter it is expected Operator 1 to be more productive. This tree (Figure 1) is a concise example of the approach proposed in this paper, which shows the great potential of such approaches for forest operations planning. For having a detail of the precision of the model statistical assessments are done. The results of these assessments are included in Table

1 as Confusion matrix table. For further research, we will study the influence of other variables on productivity and extend the analysis to other dependent variables and larger datasets.

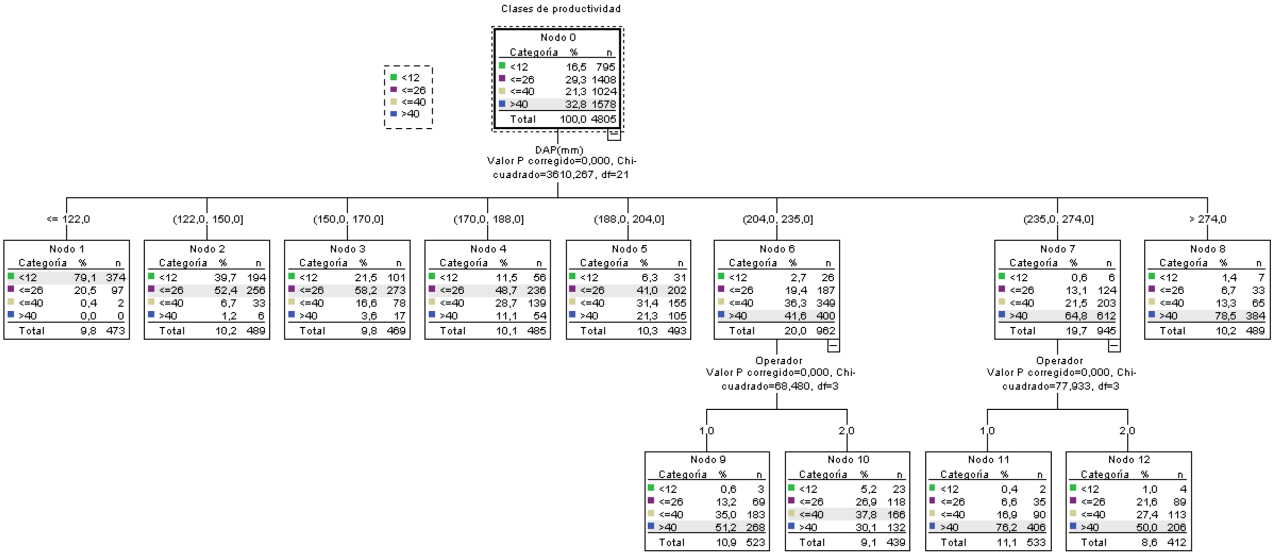


Figure 1. Decision tree model to predict the kind of productivity a harvester.

Observed	*Forecasted*				
	<12	*<=26*	*<=40*	*>40*	*Correct percentage*
< 12	374	382	23	16	47%
<=26	97	967	118	226	68.7%
<=40	2	405	166	451	16.2%
>40	0	182	132	1264	80.1%
Global percentage	9.8%	40.3%	9.1%	40.7%	57.7%

Table 1. Confusion matrix of the tree model.

Referencias

[1] Weintraub, A., y Romero, C. (2006). Operations research models and the management of agricultural and forestry resources: a review and comparison. Interfaces, 36(5), 446-457.

[2] Milanesi, G. S., Broz, D., Tohmé, F., & Rossit, D. (2014). Strategic analysis of forest investments using real option: The fuzzy pay-off model (FPOM). *Fuzzy Economic Review*, *19*(1), 33.

[3] Broz, D., Durand, G., Rossit, D., Tohmé, F., y Frutos, M. (2016). Strategic planning in a forest supply chain: a multigoal and multiproduct approach. *Canadian Journal of Forest Research*, *47*(999), 297-307.

[4] Broz, Diego Ricardo. "Técnicas de simulación y optimización aplicadas a la planificación forestal." (2015). Ediuns, Bahía Blanca, Argentina.

[5] Olivera, A., Visser, R., Acuna, M., & Morgenroth, J. (2015). Automatic GNSS-enabled harvester data collection as a tool to evaluate factors affecting harvester productivity in a *Eucalyptus spp*. harvesting operation in Uruguay. *International Journal of Forest Engineering*, *27*(1), 15-28.

[6] Ahlemeyer-Stubbe, A., Coleman, S. *A practical guide to data mining for business and industry*, John Wiley & Sons, 2014.

[7] Olivera Farias, Ernesto Alejandro. "Exploring opportunities for the integration of GNSS with forest harvester data to improve forest management." (2016).

Business Intelligence technologies for the automation and analysis of meteorological parameters for agriculture in Ancash-Peru

Rocío Rocha[1], Jesús E. Espinola[2], Ángel Cobo[1], Rafael Figueroa[2], Lluis M. Plá[3], Maximiliano E. Asís[2]

[1] University of Cantabria
Dpt. Business Administration, Dpt. Applied Mathematics and Computer Science,
39005 Santander (Spain)
rochar@unican.es, acobo@unican.es

[2] Santiago Antúnez de Mayolo University (UNASAM)
Faculty of Science, Faculty of Environmental Sciences, Faculty of Science
Huaraz (Perú)
espinolj@gmail.com, figueroatauquino.rafaelramon2@gmail.com, maxasis@hotmail.com

[3] University of Lleida (UdL)
Department of Mathematics
25003 Lleida (Spain)
lmpla@matematica.udl.cat

Abstract

The present work shows a Business Intelligence platform that integrates data mining model for the intelligent analysis of the meteorological parameters in the region of Ancash-Peru that can be used to plan the agricultural activities of the region as time of planting and harvesting. The platform allows a real-time visualization of essential meteorological parameters, such as dew point temperature, air temperature, rainfall, solar radiation, wind speed, among others. The BI platform aims to arm farmers and agricultural businesses with the data they need to make better and more informed decisions.

1 Introduction

Ancash is a region located in the central-western region of Peru. It has an area of 35.915 km^2 and its orography is very diverse, places in the region vary from zero up to 6.768 meters above sea level. Ancash`s population, according to the last census carried out in 2007 and registered by INEI [2], was 1,063,459 inhabitants. The main economic activities are mining, agriculture and tourism. Agriculture sector plays an important role in the economic development of the region. The sector employs the majority of the economically active population; however, it has not reached the required level of automation and much of the production is based on small farms or agricultural cooperatives. The region faces the effects of an inadequate agricultural practice in the management of soil, water and agrochemicals. Therefore, the developments in this sector, the introduction of new technologies and the intelligent use of available data surely help to develop the region technically as well as economically. Different Web portals, data source and services are available to help the agriculture stakeholders but the farmer community needs the information to be available in an agile way, with easy access and to integrate data from different repositories or sources. In short, the sector needs ETL (Extract, Transform and Load) tools and BI processes that facilitate the subsequent generation of useful knowledge for decision-making and increase the farmer's productivity.

In particular, it is of enormous importance to implement climate services for agriculture that enable producers to manage the risks and opportunities arising from climate variability. Although not the only one, the climate factor is one of the most important elements when planning agricultural production. Climate change and global warming is also affecting this area, and the main disadvantaged are farmers because, among other things, their usual rain and drought schedules have been altered. In this context, it is especially necessary to have reliable and focused meteorological information in the different areas of agricultural production. In this way, different agricultural activities can be properly planned.

2 Business Intelligence in agriculture

It can be a common misconception that Business Intelligence (BI) techniques and tools are tailored toward big enterprises or organizations in specific sector as finance, telecommunications, high-technology industry…, but this is not the case. All industries benefit from the use of BI techniques, one of which is the agricultural sector. Digital or data-driven farming is transforming agriculture with technology being deployed in almost every area of crop and livestock production. Ghadiyali, Lad and Patel [1] introduced the concept of Agriculture Intelligence as “a collection of integrated operational as well as decision-support applications, technologies and Databases that provide the agriculture community easy access to agriculture data.” Looking at the agricultural industry, there is a variety of uses and needs for BI solutions such as forecasted data, analysis of environmental impact, search for competitive advantages and support to decision-making.

Although precipitation is the single most important weather variable for most crop production, crop forecasts can be determined with data from additional weather data (precipitation, temperature, solar radiation, relative humidity, evapotranspiration…). On the other hand, data mining offers an interesting and useful solution to this need, as it allows obtaining trends of the different meteorological parameters in the area.

3 BI platform

In order to serve as an analytical tool, a BI platform has been implemented using different technological solutions. The platform uses ETL processes to automatically receive satellite meteorological data. The used data are obtained from a MySQL database that stores the values of the meteorological parameters sent by 16 meteorological stations (Figure 1), which are installed in different geographic points of the region. The data is received in the database via satellite, and recorded every hour. This information has been registered since 2012 and using Pentaho ETL processes, spreadsheets are generated on which predictions are obtained using statistical techniques (Figure 2). The purpose of a data mining model integrated in the platform is to provide the trends of the main meteorological parameters to predict its behavior in the following 72 hours. For this purpose, air temperature, wind speed, evaporation level from the ground, dew point temperature, precipitation, solar radiation and relative humidity have been considered as parameters of analysis. The platform tries to support the prediction of trends of the main meteorological parameters according to the information of the last days, but also it is contrasted with the tendency of the same parameters in the previous years. This allows to combine the current tendencies with the historical tendencies and to obtain trends that are more reliable. This model can run dynamically with small time intervals, allowing continuous feedback and can be applied to each zone where each meteorological station is located, which allows to have specific information, this is very important considering the varied orography of the region. The results of the trend are shown in real time through a web platform, which is very useful for farmers in the region, for time of planting and harvesting. The initial porpoise has been to have the BI platform on computers, but it could be possible to extend its use on smartphones or tablets.

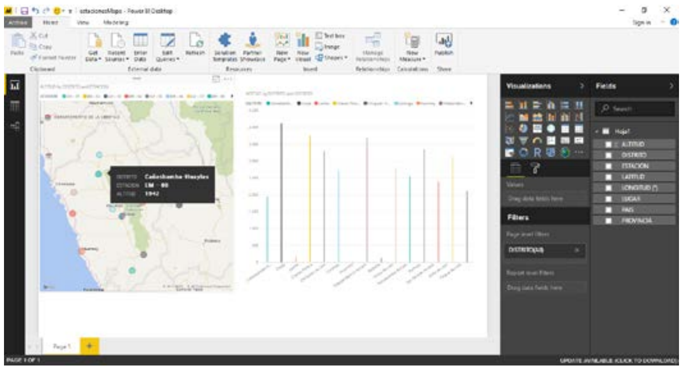


Figure 1. Visualization of meteorological stations location using powerBI.

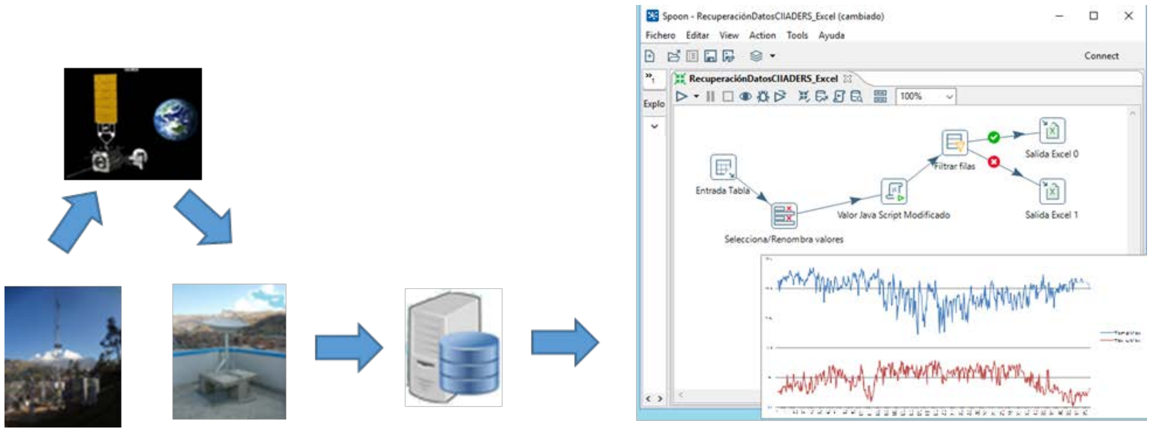


Figure 3. Sending and receiving meteorological information, ETL processes and example of temperature forecast.

Although the BI platform focuses on the integration and analysis of meteorological data, it also incorporates ETL processes to integrate data from other sources and open data portals related to the agricultural sector. In short, the platform aims to arm farmers and agricultural businesses with the data they need to make better and more informed decisions.

References

Ghadiyali, Tejas; Lad, Kalpesh and Patel, Bankim (2011). Agriculture Intelligence: An emerging tool for farmer community. *Proceedings of second international conference on “Emerging Application of Information technology”* EAIT 2011.18–20 Feb, Kolkata. 2011 IEEE DOI 10.1108\9/EAIT 2011.36, Vol.2 pp 313–316.

Instituto Nacional de Estadística e Informática (2015). INEI. Available in: https://www.inei.gob.pe/estadisticas/indice-tematico/poblacion-y-vivienda/ [last access: june 3th, 2017]

Time-dependent performance evaluation of a tire repair system in the agricultural stage of sugarcane industry

Carolina Gualberto[1], Lásara Rodrigues[1], Reinaldo Morabito[2]

[1] Federal University of Ouro Preto
Campus Moro do Cruzeiro, 35400-000 Ouro Preto, MG, Brazil
carolina_gualberto@yahoo.com.br
lasara@em.ufop.br

[2] Federal University of São Carlos
Via Washington Luiz, km 235, 13565-905 São Carlos, SP, Brazil
morabito@ufscar.br

Abstract

We propose a time-dependent hypercube queuing model and apply this model using a case study conducted in the tire repair emergency services of a sugarcane mill. The aim of this work is to analyze if the time-dependent changes and decentralization of servers have substantial impact on system's performance. The findings suggest that is important to consider the changes in the location of harvest and planting fronts and that the decentralization of servers doesn't seem to improve system performance.

1 Introduction

Studies of emergency systems applied to medical, fire, and police can be found in the literature. However, there are not many works analyzing emergency system in the agriculture stage of sugarcane industry, such as [2-4].

The hypercube model is a queuing model based on Markovian approximations for the analysis of spatially distributed queues, in which the customers and/or servers are distributed over the region. It is often used to analyze the configuration and operation of server-to-customer emergency systems, in which servers must travel to the calls locations (sites). The model was originally proposed by [1] as a specific tool for the planning and evaluation of service systems that have spatially distributed random demand.

In many queuing systems, time-dependent changes in system parameters can substantially impact on system's performance [5]. Furthermore, there are several approaches for the analysis of time-dependent queuing systems. On the other hand, to the best of our knowledge, there are no earlier studies in the literature analyzing in conjunction time-dependent parameters in spatially distributed queues through the hypercube model.

This paper provides an analytical solution for the time-dependent performance evaluation of the tire repair system in the agricultural stage of the Brazilian sugarcane industry. In this mill, the harvest and planting fronts change its locations during the harvest season. It is noteworthy that harvest and planting plans are deterministic and based on operational rules of the system. So, the location of service calls also moves leading to the arrival rate in each atom to be time-dependent.

2 The Tire Repair Emergency System

The current study involves a typical Brazilian sugar and ethanol mill, located in the State of Minas Gerais. This company farms approximately 32,000 hectares of fully irrigated sugarcane per year and harvests, transports and grinds about 2.7 million tons of sugarcane annually. The agricultural area is divided into 10 geographical atoms in this study, which represent the different sugarcane harvest areas. Furthermore, the entire harvest and planting process is mechanized, without any burning of the cane.

In this company, activities carried out in the agricultural stage use various types of equipment and vehicles, such as harvesters, planters, tractors, loading units, and trucks. And, when tire repair is needed, this equipment is unable to travel to the tire repair shop, requiring the dispatch of a tire repair truck to perform the service in the field. The company's tire repair structure includes two tire repair trucks, the tire repair shop (in atom 4) and the support facility (near the mill in atom 1). The trucks are operated by tire repairmen teams that work in three daily shifts.

In the original scenario, the trucks remain fixed and centralized in the tire repair shop's support facility, and when necessary, travel to emergency sites. The servers can be considered homogeneous, in that they all had the same type of truck, location, and random dispatch rule (since there is no server-based dispatch preference).They may travel to any part of the fields within the agricultural stage; and when the service is completed, they are directed back to their base when no further service calls waiting in queue exist. If there is an outstanding call, the server goes directly to that site, without having to go back to its base.

Additionally, the tire repair truck's decentralization is under study by the company managers and an alternative configuration is also analyzed in this work. In this alternative scenario, server 1 remains located in atom 1 while server 2 is at atom 4, when available. Note that with the two servers decentralized, different dispatching policies are defined for each server in a fixed dispatch preference list. In this case, the servers should be distinguishable in the analysis, while sharing the demand of all atoms, because of their different dispatching policies depending on their locations and the atom locations.

For additional details regarding average travel times between the atoms, service rates and dispatch preference list in alternative scenario see [2, 3]. With respect to the arrival rates, we used realistic data based on the current operation of the tire repair system, since data regarding changes in front locations are not fully available.

Previous studies of the tire repair system [2, 3] analyzed the system assuming the steady state assumption. It seems to be reasonable looking at the average arrival rate. However, as the harvest and planting fronts change its locations during the harvest season, the location of service calls also moves and should be considered. Due to this, in the current study, we analyze the time-dependent behavior of this system under non-stationary conditions.

The classic Hypercube Model [1] was adapted to describe the dynamic behavior of this system. Both original and alternative scenarios were analyzed assuming the system is non-stationary due to time-dependent changes in arrival rates in each atom. The set of differential equations for the original scenarios is shown in (1). As in the tire repair system, there are two servers and 4 (2^n) states associated with the hypercube, in which n is the number of servers, as follows: $\{00\}$, $\{01\}$, $\{10\}$ and $\{11\}$. These states represent the possible combinations associated with each server: free (0) or busy (1). In systems that use total backup, the queue only arises from state $\{11\}$, and is denoted by states $\{S_m\}$, in which m is the number of call in the system. The total arrival rate of the tire repair service in each time interval is then $\lambda(t) = \sum_{i=1}^{10} \lambda_i(t)$, while the total service rate is $\mu = \sum_{j=1}^{2} \mu_j$, i.e., the service rates are constant during the entire time interval. As the alternative scenario applies a dispatch preference list, the set of differential equations is different for states $\{10\}$ and $\{01\}$ compared to (1).

(1)

3 Results

The time-dependent hypercube model was applied to analyze the original and alternative configurations of the mill. The sets of differential equations for both scenarios were solved by the Matlab function ode45 to calculate the state probabilities in each time interval.

The workload of each server i ($\rho_i(t)$) in both scenarios is shown in Figure 1. In the original scenario, as the servers are centralized and assumed homogeneous, the workload of servers 1 and 2 are the same. On the other hand, the workload in the alternative scenario reflects the dispatch policy of the system and the average of servers 1 and 2 is equal to the workload in original scenario.
Regarding the number in queue and in the system, the maximum deviation between original and alternative scenarios reached 0.56%.

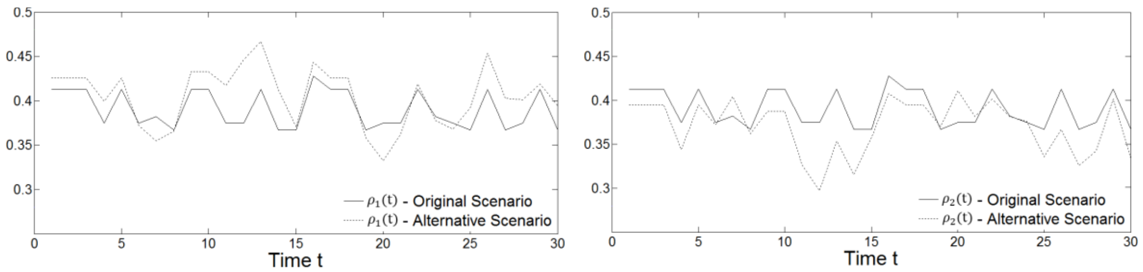


Figure 1: Server workloads

4 Concluding remarks

In this paper, we developed an approach for the time-dependent performance analysis of the tire repair systems in sugarcane mills. The result suggests that is important to consider the changes in the location of harvest and planting fronts. Regarding the scenarios, our findings show that decentralization of servers in alternative scenario doesn't seem to improve the system performance. With respect to future research, we intend to evaluate other alternative scenarios, dispatch policies, performance measures and solution approaches.

Acknowledgements

The authors would like to thank the Sugarcane Mills for its collaboration with this research, and FAPEMIG, CNPq, UFOP and Fundação Gorceix in Brazil for their financial support.

References

[1] Richard Larson. Hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, 1: 67-95, 1974.

[2] Lásara Rodrigues. Análise dos Serviços Emergenciais de Manutenção Agrícola e Borracharia na Agroindústria Canavieira utilizando Teoria de Filas. PhD Dissertation. Universidade Federal de São Carlos, São Carlos, Brazil, 2014.

[3] Lásara Rodrigues, Reinaldo Morabito and Cem Saydam. Queuing Analysis of Emergency Repair Systems: a Case Study of Tire Repairing in Brazil Sugarcane Industry. *International Journal of Operations and Quantitative Management*, 22(4): 335-356, 2016.

[4] Lásara Rodrigues, Reinaldo Morabito, Fernando Chiyoshi, Ana Iannoni and Cem Saydam. Towards hypercube queuing models for dispatch policies with priority in queue and partial backup. *Computers & Operations Research*, 84: 92-105, 2017.

[5] Justus Schwarz, Gregor Selinka and Raik Stolletz. Performance analysis of time-dependent queueing systems: Survey and classification. *Omega*, 63: 170-189, 2016.

Game theory concepts and changes in the Brazilian agriculture

Fernando L. Garagorry

Embrapa, Secretaria de Gestão e Desenvolvimento Institucional
Ed. Sede, Caixa Postal 08815, 70770-901 Brasília, DF, Brasil
fernando.garagorry@embrapa.br

Abstract

The evolution of the Brazilian agriculture, over the last four decades, has shown substantial changes in the geographical distribution of individual products, land use, production value and many other variables. Besides, even within a fixed geographical area, many changes have taken place. The study of these variations is based on large amounts of available data and the use of non-traditional analytical techniques, in order to obtain new information. One of these techniques, based on the power indices of game theory, has been applied to datasets on individual products, land use and production value.

1 Introduction

The Brazilian agriculture has shown, over recent decades, substantial changes in the geographical distribution of products, land use, manpower, machinery, production values and many other variables. The study of this kind of motion over the national territory is being conducted under the general name of "agrodynamics" [2]. Following the official Territorial Division of Brazil, the studies have considered the levels of the whole country (1), regions (5), states (27), mesoregions (137) and microregions (558), for a total of 728 geographical entities. Nevertheless, many data are available down to the level of municipalities (their number has changed over the years, and presently is 5570). Besides the changes in geographical terms, other variations have taken place within a fixed territorial entity. In any case, many types of changes can be identified and assessed with the same techniques.

The following situations have been considered: a) distribution of the "volume" (i.e., stock head-count for animals or produced quantity for all the other items) of individual products in the five regions, with annual data, many going back to 1975; b) distribution of agricultural area into six types of land use; c) distribution of monetary value of agricultural production into eigth classes. In the last two cases, the data come from six agricultural censuses. In all these situations, the main analytical techniques are based on a simple step: in any year, the original values in the different classes are divided by the total of the respective additive variable and a relative distribution is obtained. In the paper, the numbers are multiplied by 100, so that the situations are presented as sets of percent distributions.

Since a distribution can also be seen as a voting game, it is possible to determine some indices of power for that game. Here, only the Banzhaf-Coleman and the Shapley indices will be considered. There are some important things to be mentioned about these indices: 1) they have been presented as "solution concepts" for cooperative games; 2) the same index of power can correspond to different games; 3) usually, the indices of Banzhaf-Coleman and of Shapley are not very different; 4) the Banzhaf-Coleman index is easier to evaluate than that of Shapley. The main idea is to use such indices to discriminate between seemingly close distributions or to group seemingly quite different ones. That is, the use of power indices appears as an auxiliary data mining technique.

2 Indices of power

An introduction to the theory of n-person cooperative games can be found in Owen [3, Chapters 8-11]. We consider an n-person game, with $n > 2$, which has the form of a simple majority voting game. N will denote the set of players. Throughout this paper, the original values of the players will be presented as an ordered set of n non-negative numbers, which add up to 100. The value of a coalition is the sum of the values of its members. A winning coalition is anyone whose value is greater than 50. A *swing* for player i is defined as a set $S \subseteq N$ such that $i \in S$, S wins, and $S - \{i\}$ loses.

Letting b_i be the number of swings for player i, the following number can be calculated:

. (1)

This is the normalized Banzhaf-Coleman index for player i. More generally, the Banzhaf-Coleman index of power, for the original game, is the list $B = (B_1, B_2, \cdots, B_n)$. By construction, this is a list of non-negative numbers which add up to *1*; in this presentation, these numbers will be multiplied by *100*. For simplicity, the list will be referred to as the Banzhaf index of power. A detailed discussion of the mathematical properties of this index was presented in Dubey and Shapley [1].

In a similar way, at least for the type of games being considered here, the Shapley value for player i can be defined as follows:

———— , (2)

where the summation is taken over all swings T for player i, and t is the number of players in T. The Shapley value for the original game is the list $S = (S_1, S_2, \cdots, S_n)$. It is well known that this is a list of non-negative numbers which add up to *1*; in this presentation, as before, these numbers will be multiplied by *100*.

Therefore, each of the above lists, that is the Banzhaf index or the Shapley value, can be taken as defining a new simple majority game, which was deduced from the original game. The important thing is that they capture some essential facts relating to the winning coalitions of the original game. For some games both lists coincide, but this is not always the case.

3 L1 distance between two distributions

As presented above, a game can also be seen as a percent distribution; that is, a set of non-negative numbers which add up to *100*. Given two percent distributions, as ordered lists with n components, corresponding to years s and t, such as $f^s = (f_1^s, f_2^s, \cdots, f_n^s)$ and $f^t = (f_1^t, f_2^t, \cdots, f_n^t)$, the L1 distance between them will be defined as

. (3)

The factor ½ is used so that the distance takes values between *0* and *100*. This distance will be applied to the original games, as well as to the respective power indices.

4 Results

The results will be illustrated with the example of rice. The annual data, starting in 1975, are available down to the level of municipality. Here, only regional distributions will be shown. It is well known that a motion to the South region has taken place over the years. But a more detailed analysis is required. Table 1 shows a typical situation where power indices can be used to detect changes in the structure of winning coalitions. For any pair of consecutive years, the distance appears on the second line of the pair. For convenience, the distributions corresponding to the original production data are labelled as type "DATA"; the others correspond to the respective power indices.

Table 1 shows rather small distances for the data, and much larger values for the power indices. In particular, from 1989 to 1990, the percent contribution of rice in the South region changed from a value smaller than *50* to a value larger than that. When the data distribution for 1990 is seen as a voting game, it has a dominant player and all the others are dummies. In that kind of situation, the indices of power give a value of *100* to the dominant player and *0* to the others.

Figure 1 illustrates a typical result with regard to the three sets of distances. In several cases, small changes in the distance between the original distributions coincide with abrupt changes in those calculated with the power indices.

Type	Year	North	Northeast	Southeast	South	Center-West	Distance
DATA	1988	9.73	17.62	13.60	40.23	18.82	–
	1989	10.63	16.18	13.18	43.63	16.38	4.30
	1990	8.16	11.53	13.87	54.11	12.33	11.17
BANZHAF	1988	0.00	16.67	16.67	50.00	16.67	
	1989	9.09	9.09	9.09	63.64	9.09	22.73
	1990	0.00	0.00	0.00	100.00	0.00	36.36
SHAPLEY	1988	0.00	16.67	16.67	50.00	16.67	
	1989	10.00	10.00	10.00	60.00	10.00	20.00
	1990	0.00	0.00	0.00	100.00	0.00	40.00

Table 1: L1 distances for sets of distributions related to rice production quantity

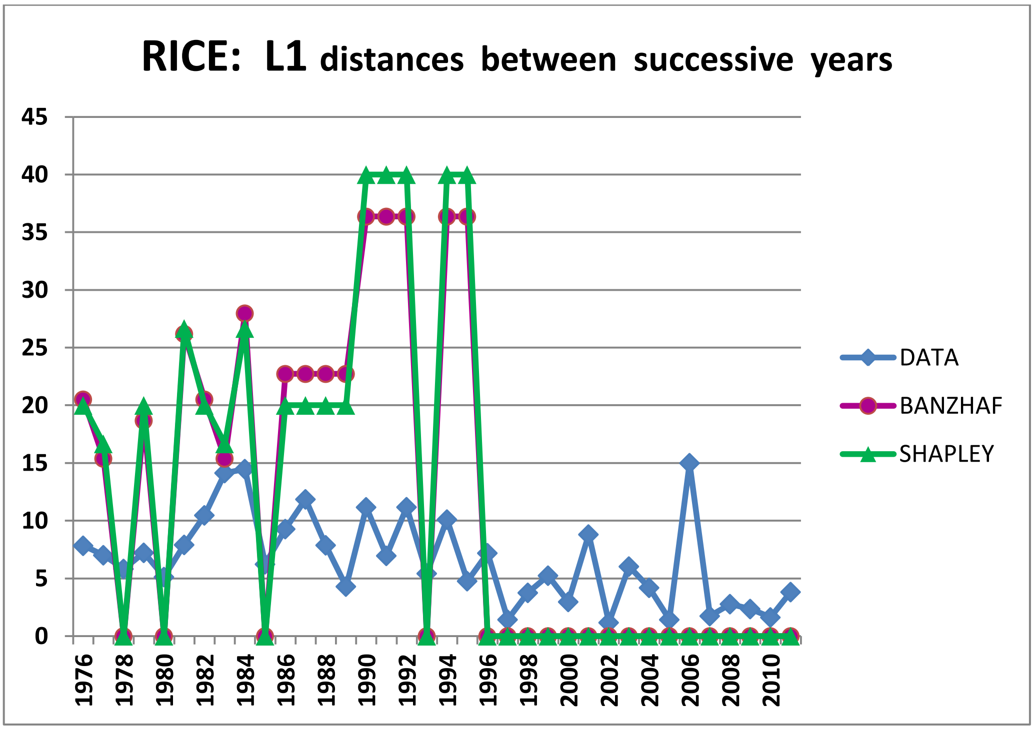


Figure 1: L1 distances, for data and power indices, in successive years

In such cases, the distances between the indices show greater discriminating power than those derived from the original distributions. For that reason, power indices are being used as an auxiliary data mining technique.

References

[1] Dubey, P. and Shapley, L. S. Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research*, 4:99-131, 1979.

[2] Garagorry, F. L. and Chaib Filho, H. Concentration and dynamics of the Brazilian agriculture. In Prado, H. A. do, Luiz, A. J. B. and Chaib Filho, H., editors, *Computational methods for agricultural research: advances and applications*. Hershey, Information Science Reference, 2011.

[3] Owen, G. *Game theory*. New York, Academic Press, 1982. 2nd. ed.

A Stochastic Frontier Approach in the Presence of Endogeneity for the Brazilian Agriculture

Geraldo da Silva e Souza[1], Eliane Gonçalves Gomes[2]

[1, 2] Embrapa, Secretaria de Gestão e Desenvolvimento Institucional
Parque Estação Biológica, Av. W3 Norte final, 70770-901, Brasília, DF, Brazil
[1] geraldo.souza@embrapa.br
[2] eliane.gomes@embrapa.br

Abstract

Using farmers' data comprising the Brazilian agricultural census of 2006, we fit a stochastic frontier model in the presence of endogeneity. The endogenous variable affecting efficiency is the probability of negative net income. Additionally we investigate the effects of technical assistance and agricultural research in efficiency. Production is modeled in the Cobb-Douglas family with three inputs: expenditures on labor, on land, and on other technical inputs. The response is gross income. Technological inputs are dominant, followed by labor and land. Research and technical assistance contribute significantly to reduce inefficiency and the probability of negative net income contributes positively to inefficiency.

1 Introduction

The objective of this article is to fit a production model for the Brazilian agriculture using census microdata (information at the farm level). Our analysis refines the work of [3] and aims at investigating the relative size of input elasticities of labor, land and capital, and the efficiency effects of agricultural research, technical assistance and the probability of negative net income. The latter is viewed as an endogenous variable.

The Brazilian agricultural census of 2006 inquired 5,175,636 farmers. A total of 2,039 variables were measured. We took a random sample of observations in order to accommodate a complex regression analysis on this dataset. The sampling strategy adopted followed a stratified random sampling scheme with proportional allocation. The sample size was fixed assuming a maximum error of 150.00 BRL in the estimation of the mean gross farm income, with probability of 95%, with known strata variances. The strata were based on both regional and gross income strata [3]. The reference for defining income output brackets was the minimum wage in force in Brazil. For 2006 this value was 300.00 BRL. The combined stratification of (5) regions and (4) income output brackets led to a total of 16 strata. The last stratum included rich farmers (i.e. with high income values), which were grouped together regardless of the region. Income brackets were defined as follows, across the five Brazilian regions (North, Northeast, South, Southeast and Center-West): A - annual gross income in the range (0, 7,200.00], B - annual gross income in the range (7,200.00, 36,000.00], C - (36,000.00, 720,000.00]. Farmers were considered rich if their income exceeded 720,000.00 BRL. The sample contained 258,684 farms out of a population of 4,614,030 farms with valid data. The need for proper instrumental variables reduced the sample size to 74,149 observations.

2 Stochastic frontier

The analysis of a stochastic production frontier model was conducted on the sample (74,149 farms). The frontier was taken from the Cobb-Douglas family and the model is given by (1):

$$\ln(y_j) = \beta_0 + \beta_1 \ln(xtrab_j) + \beta_2 \ln(xterra_j) + \beta_3 \ln(xtec_j) + \beta_4 D_{1j} + \beta_5 D_{2j} + \beta_6 D_{3j} + \beta_7 D_{4j} + v_j - u_j \quad (1)$$

where ln is the natural log, y represents gross income, *xtrab* are expenditures with labor, *xterra* are expenditures with land and *xtec* are expenditures with technological inputs for the j producer. $D1$ to $D4$ are dummy variables for four regions, whose coefficients represent differential effects relative to the intercept, which represents the Center-Western region. The components vj and uj represent the random error and the inefficiency component of the model. The normal-half-normal combination for these components was chosen as it best fits the data. The model is heteroskedastic.

The random error component (*vj*) has variance dependent on stratum dummies. The variance of the inefficiency component (*uj*) is explained by contextual variables. Here, we chose the following variables: probability of negative net income (*p*), agricultural research efforts (*score*), access to technical assistance (*assitec*) and regional dummies. The variable *score* measures, on a qualitative scale (low, average, high), the experts' perceptions of the influence of Brazilian Agricultural Research Corporation (Embrapa) on agriculture. The probability of negative net income was estimated in a first-stage via a probit regression (results are not reported here). This was done to control for endogeneity in the income (output) equation. We therefore resorted to an instrumental variable approach, which regressed the dichotomous dummy variable indicating whether or not the farm had a negative net output against a set of exogenous variables. These variables included the inputs described above and the following variables, which in addition to serving as instruments are also relevant in explaining changes in technical inefficiency: farmer's experience, type of production (whether crops, livestock or mixed), education of the head of household, age of the head of household, family size, access to cooperatives, whether the farmer was in urban or rural setting, gender of the head of household, access to credit, access to technical assistance, research efforts and whether the farmer rented or owned his farm. Then, fitted values of the estimated probabilities were added to the regression outlined above and the final set of instruments included, additionally, the stratum dummies. The model was estimated by maximum likelihood to account for endogeneity bias, as proposed by [1, 2].

3 Statistical results

Table 1 shows the results of our econometric exercise. It is important to point out here that stratum effects are fixed since strata are defined by the sampling plan. The model though has two variance components, associated with the idiosyncratic error and with the inefficiency random error. We omitted from Table1 the estimates of the stratum effects on the idiosyncratic variance component due to lack of space, but they jointly affect significantly the corresponding variance component.

All relevant effects for explaining the variation in technical inefficiency are significant with the proper signs, indicating that technical inefficiency decreases monotonically as a function of technical assistance and research efforts, and increases as a function of *p*. The regional efficiencies are in the following order (higher to lower): Center-West, Southeast, South, North and Northeast. Land elasticity presented in Tables 1 and 2 is considerably lower than elasticities of labor and especially technological inputs. This result has strong implications for technology diffusion and suggests that farmers that do not adopt technological inputs face a dire predicament in terms of output.

4 Concluding remarks

Some key results are worth emphasizing. To start, the relative size of the estimated elasticities indicates the dominance of technological inputs over other inputs. Technical assistance, along with agricultural research, contributes to the reduction of technical inefficiencies.

The predicted probabilities obtained from the probit regression, which instrumented the dummy variable indicating whether or not farmers had a negative net output, was significant and of extreme importance in reducing inefficiencies. While this finding is likely to be obvious, it highlights the importance of controlling for farmers that are not capable of generating a sufficient or at least non-negative net output.

Technical assistance had a positive effect on technical efficiency. At the same time, negative net outputs are strongly associated with inefficiency. This suggests that the technical assistance being provided is outweighed by the lack of financial and managerial skills of farmers. These skills have thus far not been provided by technology through technical assistance.

Technology is knowledge created by research and applied by producers through production systems. Thus, it seems evident that only a selected few larger farmers were able to fruitfully develop production systems that benefit from technology. Small scale agriculture needs to be reassessed and refocused to be able access technology and become profitable.

This work points out that productive inclusion in Brazil can be achieved only by giving access to technology to the small farmers (low income). Asymmetries in credit, infrastructure and education create barriers to rural extension and to the visibility of agricultural research results. Public policies should therefore be oriented to reduce the existing asymmetries.

Finally, it is important to stress the article's contribution to the estimation of stochastic frontier models under endogeneity. The refinement in estimation proposed here is new in the context of applications in the agroeconomics literature.

	Coefficient	**Standard Error**	**p-value**
l*y*			
l*xtrab*	0.2160	0.0032	<0.001
l*xterra*	0.0873	0.0022	<0.001
l*xtec*	0.6297	0.0037	<0.001
North	0.1992	0.0330	<0.001
Northeast	0.0937	0.0252	<0.001
Southeast	0.0962	0.0242	<0.001
South	-0.0652	0.0227	0.004
constant	2.5868	0.0366	<0.001
lnsig2u			
p	6.6334	0.0497	<0.001
score	-0.0555	0.0209	0.008
assitec	-0.0477	0.0242	0.048
North	0.4058	0.0601	<0.001
Northeast	0.5797	0.0483	<0.001
Southeast	0.1017	0.0506	0.005
South	0.1454	0.0450	0.004
constant	-4.1718	0.0778	<0.001

Table 1: Production function estimation results.

Inputs	**Elasticities**	**Share**
Labor	0.2160	23.15
Land	0.0873	09.36
Capital	0.6297	67.49

Table 2: Inputs' contributions.

References

[1] Mustafa Karakaplan. Fitting endogenous stochastic frontier models in Stata. *The Stata Journal*, 17(1): 39–55, 2017.

[2] Mustafa Karakaplan and Levent Kutlu. *Handling Endogeneity in Stochastic Frontier Analysis*. 2013. http://www.mukarakaplan.com/Karakaplan - EndoSFA.pdf

[3] Geraldo Souza, Eliseu Alves, Eliane Gomes, Eduardo Magalhães and Daniela Rocha. Um modelo de produção para a agricultura brasileira e a importância da pesquisa da Embrapa. In: Eliseu Alves, Geraldo Souza and Eliane Gomes, editors, *Contribuição da Embrapa para o Desenvolvimento da Agricultura no Brasil*, pages 47–86, 2013.

Geostatistical study of root rot produced by the fungus *Rhizoctonia solani* Kühn in the cultivation of *Vigna unguiculata* (L.) Walp in the municipality of Gibara, Cuba

Vilma López[1], Esteban López[2], José Q. Cuador[3], Ramón C. Núñez[1]

[1] University of Holguin
CEAAR, University of Holguin, Av. XX Aniversario s/n, 80100 Holguín, Cuba
vlopez@uho.edu.cu, rnunez@uho.edu.cu

[2] University of Holguin
Mechanical Engineering Department, University of Holguin, Av. XX Aniversario s/n, 80100 Holguín, Cuba
elopez@uho.edu.cu

[3] University of Pinar del Rio
Departamento de Matemáticas, University of Pinar del Rio, A4, 20100 Pinar del Río, Cuba
cuador@upr.edu.cu

Abstract

In this article the use of geostatistical techniques are shown to analyze the spatial variability of the incidence of root rot in growing *Vigna unguiculata* (L.) Walp, produced by the fungus *Rhizoctonia solani* Kühn. An initial exploratory analysis was performed on the disease data, which guaranteed the applicability of geostatistical techniques. Using the SURFER software, the experimental and theoretical variograms were determined for each sampling. With the theoretical models of variograms the analysis of the spatial structures was made. The obtained maps allowed to observe the spatial distribution of the disease and its evolution over time.

1 Introduction

The specie *Vigna unguiculata* (L.) Walp, known in Cuba as *Caupí* bean, *Cancarro* bean or *Carita* bean, suffers the attack of numerous pests and diseases such as the root rot caused by the fungus *Rhizoctonia solani* Kühn.

Several authors point out that it is in the initial stages of cultivation that major affectations produced by soil fungus are present, especially because of this pathogen [3, 4, 6]. On the other hand, another author [1] states that *R. solani* Kühn survives in the soil for several years; in the organic matter of the previous crop, or in the dead organic matter present in the soil.

At present, although the distribution of *R. solani* Kühn in the cultivation of cowpea is known, there is no information on the structure of the spatial distribution of *R. solani* Kühn in *Vigna unguiculata* (L.) Walp.

The present work is the first report of the use of geostatistical tools to determine the spatial distribution of the infestation of this crop, caused by the fungus *R. solani* Kühn.

2 Materials and methods

2.1 Selection of experimental area and experiment design

The field experiments were carried out at the CPA "Augusto César Sandino", from Gibara municipality, Holguín province, Cuba. An area was selected so that its geographical location was representative, it had six hectares, with soil of the type “Fersialítico Pardo Rojizo Ócrico Ferromagnesial” according to the "New Genetic Classification of Soils of Cuba".

Seven evaluations of the disease were carried out. The evaluations were doing at 15, 22, 28, 35, 42, 49 and 56 days of cultivation. Around each sub-point five random plants were taken to determine those exhibiting symptoms characteristic of *R. solani* Kühn. In each sampling, 20 plants per point were collected, 120 plants per hectare, which are equivalent to 720 plants. At the end of the evaluations, 5040 plants were analyzed in the laboratory, which were necessarily uprooted.

The experimental design scheme for each of the selected hectares is shown in Figure 1, where each grid is a sampling point (27 x 42 m). Four sub-points were established in each point, where the plant samples were taken for evaluation.

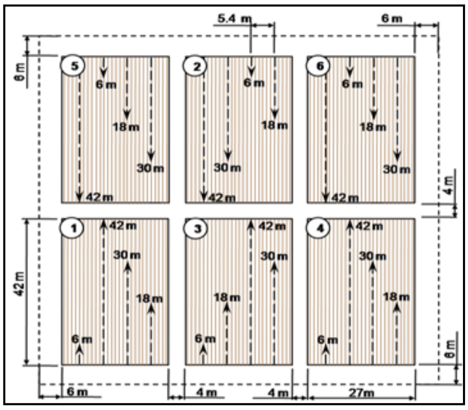


Figure 1. Design of the experimental area.

2.2 Geostatistical analysis

The database for geostatistical processing was drawn with the coordinates of the sampling points and the number of plants affected by *R. solani* Kühn in each of these points, in the seven evaluations. This information was tabulated in an ASCII file and processed with the SURFER 8 program, so that the experimental and theoretical variograms were determined for each sampling and the analysis of the spatial structures was performed. Figure 2 shows the variograms obtained at 15, 35 and 56 days.

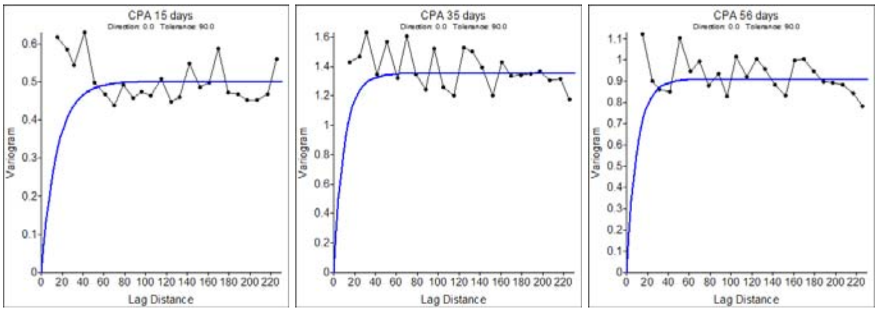


Figure 2. Experimental and theoretical variograms at 15, 35 and 56 days.

The theoretical variograms were fitted to the exponential model. It could be one or several factors originating this distribution, such as: the seed used for sowing was not in its top quality (infested seed) and its distribution in the fields was not uniform, the existence of areas with a higher moisture content than others, and the existence of pathogen inoculum sources in the soil belonging to the previous crops.

After the fitted models, the distribution disease maps were drawn for each evaluation. The resulting maps allowed to observe the spatial distribution of the disease and its evolution over time. Figure 3 shows the maps obtained to 15, 35 and 56 days. The knowledge of the spatial distribution of the disease, and the determination of its distribution patterns has made it possible to develop efficient sampling plans that provide the necessary elements for the decision making on disease control.

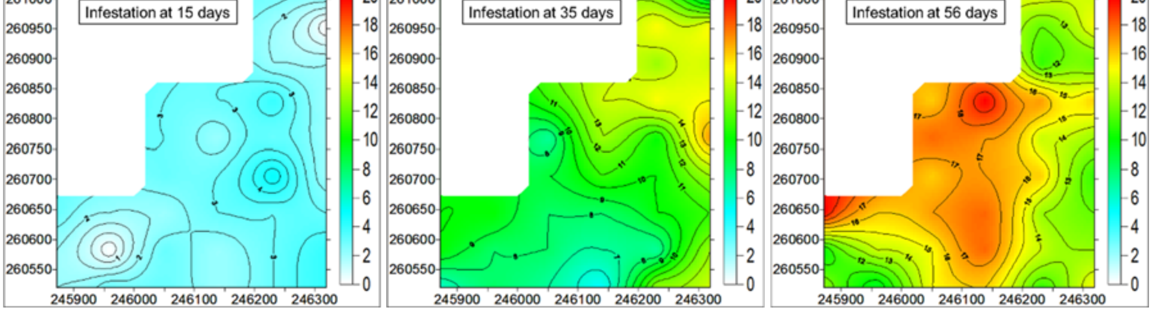


Figure 3. The distribution maps obtained at 15, 35 and 56 days.

2.3 Evaluation of the disease

When assessing the incidence of the disease, it should be noted that the selected area had been left fallow for a period of three years, due to high infestations caused by soil pathologies such as *R. solani* Kühn. This situation must have favored the presence of the fungus in the early stages of cultivation.

The required moisture of the crop in its initial stage, and the probable use of seeds infested by *R. solani* Kühn, also favor the presence of this fungus in the soil and its manifestation in the infestation of the plants.

The study showed the high level of spatial dependence between the data of each evaluation and indicated an aggregate distribution of *R. solani* Kühn in the *Caupí* bean crop in the evaluated area. The disease was present in all evaluations. From the initial stages of cultivation (15, 21 and 28 days), high pathogen infestation zones with values from 4 to18 affected plants were detected, mainly located at the edges of the plots.

In the carried out evaluations at 35, 42, 49 and 56 days, corresponding to the flowering stage (30 to 40 days), the pod formation stage (40 to 50 days) and the pod filling stage (50 to 60 days), pathogen infestation zones with values from 6 to 19 affected plants are observed. As the crop cycle advanced, the infestation occupied a larger area and spread (or moved) to other parts of the evaluated area.

The obtained results coincide with other studies of the incidence of *R. solani* Kühn in *Phaseolus vulgaris* and *Vigna unguiculata* [2, 5].

3 Conclusions

The current paper is the first one to be carried out in Cuba, using geostatistical tools, to characterize the spatial distribution of the disease caused by *R. solani* Kühn in the species *Vigna unguiculata*, which provides scientific support for alternative efficient and sustainable integrated control.

It is demonstrated that the disease can be caused by the presence of this fungus in the soil. The required moisture of the crop in its initial stage, and the probable use of seeds infested by *R. solani* Kühn, also favor the presence of this fungus in the soil and its manifestation in the infestation of the plants.

The study of the spatial distribution of *R. solani* Kühn, in growing *Vigna unguiculata* L Walp., will allow to develop a methodology for characterizing the structure and interpretation of the spatial distribution of this pathogen. It will facilitate decision-making and the implementation of planned and effective phytosanitary controls in specific areas of infestation.

References

[1] D. Henderson. Trasplantes y Plantas. On line: http://ceimperial.ucdavis.edu/newsletterfiles/Articulos_Elejidos_de_Agricultura28523.pdf. 2010.

[2] E. Pupo, A. Polanco; V. López. Micoflora asociada a semillas de Caupí (Vigna unguiculata L. Walp.) de producción local. IX Jornada Científica Productiva sobre Vigna en Cuba – Taller Internacional de Vigna en el Trópico, Holguín, Cuba, 70p. 2003.

[3] G. N Agrios. Fitopatología. Tercera edición. Editorial Limusa. México, D.F. 2005.

[4] Instituto Colombiano de Agricultura (ICA). Thanatephorus cucumeris (Frank) Donk. (Basidiomycota: Basidiomycet.,es: Ceratobasidiales: Ceratobasidiaceae). (Teleomórfico). On line: http://www.ica.gov.co/publicaciones/plagas/aster/asterendemica/aster61.htm. 2009.

[5] J. Milán, V. López, E. Fornet. Caracterización de la estructura de la distribución espacial de *Rhizoctonia solani* en el cultivo de *Vigna unguiculata*, en la Provincia de Holguín. Tesis presentada en opción al título de Máster en Producción Vegetal, Universidad de Granma, Bayamo M. N., 67p. 2012.

[6] S. Shama, H. N Raghunatham; H.S. Shetty. Seed transmition of Rhizoctonia solani in Cowpea. Geobias 21 (4-7), 79-92p. 2008.

Sugarcane Yield Estimate Analysis by using Regression Error Characteristic Curves (REC Curves)

Luiz Henrique A. Rodrigues[1], Felipe Ferreira Bocca[1]

[1] University of Campinas – School of Agricultural Engineering
Av. Cândido Rondon, 501 - Barão Geraldo 13083-875 - Campinas/SP - Brazil
lique@g.unicamp.br
felipe.bocca@feagri.unicamp.br

Abstract

Sugarcane yield estimate is one of the most important factors for harvest planning. Currently, experts use their knowledge, field history, and field assessments to perform yield estimates. A margin of tolerance can be defined for the estimates to classify errors as acceptable. By using data from a Sugarcane Mill in Brazil, this study presents a preliminary analysis of errors obtained by the comparison between estimated and real sugarcane yields using Regression Error Characteristic Curves. Results showed that in some conditions, estimates are worse than in others and indicates that further studies can identify error patterns to improve those visual estimates.

1 Introduction

Sugarcane is of great economic importance to Brazil for supplying the raw material for the production of sugar and alcohol. This production depends on the amount of raw material available, a factor of great importance for the planning of agricultural activities, especially all transportation logistics for the harvesting. There are many techniques for estimating sugarcane yield such as remote sensing [1] and soil fertility assessment [2]. Estimates made just before the harvesting season are often carried out by experts who survey sugarcane plantations observing the sugarcane growth, adding their knowledge about cultivated area and the results recorded from previous years [3,4]. This makes the results subjective and subject to significant errors. Errors in estimates impact the entire planning of sugarcane production, from harvesting to processing [5]. The identification of error patterns from the estimates will allow adjustments that will lead to better results in estimates of subsequent years. The aim of this paper is to present preliminary analyses of the errors obtained by the comparison between estimated and real sugarcane yields by using Regression Error Characteristic (REC) Curves [6]. REC Curves plot error tolerance in the x-axis and percent of points predicted within the tolerance in the y-axis; they estimate the cumulative distribution function of the error (absolute difference between predicted and real value).

2 Materials and Methods

2.1 Data

Sugarcane production data sets were provided by Odebrecht Agroindustrial from mill 'Unidade Santa Luzia (USL)', located at Nova Alvorada do Sul, MS, Brazil, for the harvest seasons 2012/2013, 2013/2014, 2014/2015 and 2015/2016.

2.2 Methods

2.2.1 Data pre-processing and transformation

The original dataset was processed to be used in this study. Instances with missing data and outliers were eliminated according to the following rules: (i) variety with frequency lower than 20; (ii) no data for number of days since last harvest; (iii) estimated yield equal to real yield; (iv) error higher than 75 $ton.ha^{-1}$. The final dataset (7943 instances) has the following attributes: harvest season (year), the number of harvests (cut), sugarcane variety, estimated yield (estimated tons of cane per hectare) and actual yield (tons of cane per hectare actually harvested). "Error" was calculated by the difference between the estimated and actual yield. Table 1 presents the distribution of the dataset concerning cut number.

	Cut #1	Cut #2	Cut #3	Cut #4	Cut #5
%	23.3%	20.8%	22.3%	20.2%	13.3%
# instances	1854	1656	1770	1605	1058

Table 1: Distribution of data concerning Cut number

2.2.2 Regression Error Characteristic (REC) curves

REC curves were created for (i) each harvest season, (ii) most frequent sugarcane varieties and (iii) number of harvests (cut number) to make analysis by comparing them with the full REC curve, made with the complete data set.

3 Results

The main analyses carried out in this work were related to three attributes: the harvest season, the sugarcane variety and the number of harvests (cut number).

3.1 Analysis for each harvest season

By analyzing the REC curves for each harvest season, we could observe that, in general, up to an absolute error around 5 t ha^{-1}, estimates were similar in all years. When the absolute error tolerance increased (x-axis), the worst estimate occurred in 2015/2016, according to Figure 1.

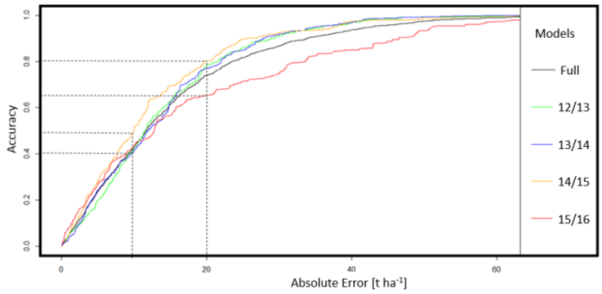


Figure 1: REC Curves for each harvest season

3.2 Analysis for most frequent sugarcane varieties

Sugarcane varieties RB867515 (38%), RB855453 (11%) and SP83-2847 (11%) were predominant. By analyzing the REC curves for those varieties, we could observe that, in general, estimates were similar and there is no difference among them, according to Figure 2.

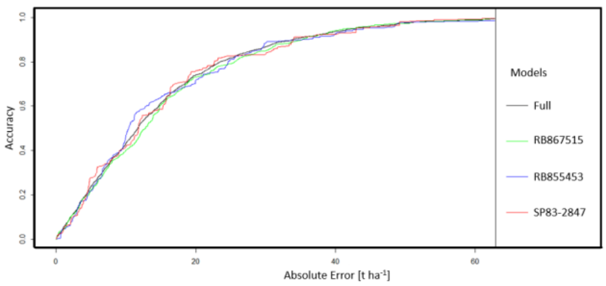


Figure 2: REC Curves for each sugarcane variety

3.3 Analysis for number of harvests (cut number)

For the analysis of REC curves for each cut number, 1st cut was analyzed separately. Plant cane, as this is called, usually has higher yields and 1st cut can be done 12, 15 or 18 months after planting. By analyzing the REC curves for each cut number, we could observe that, in general, estimates for the 1st cut were worse than estimates for 2nd cut onwards, according to Figure 3.

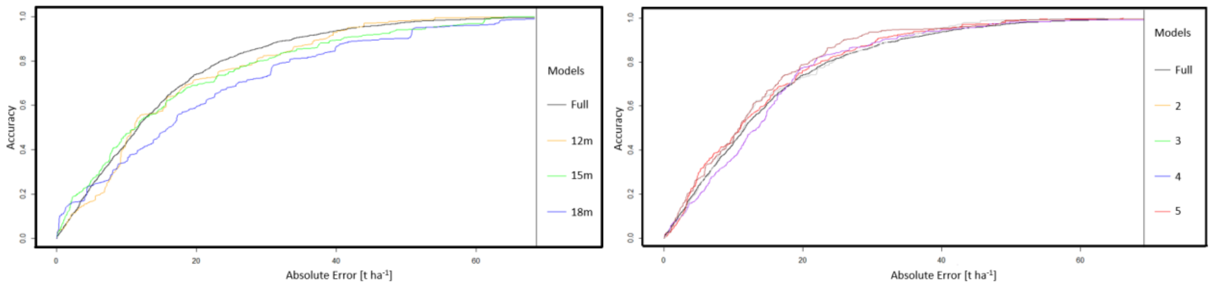


Figure 3: REC Curves for 1st cut number (left) and 2nd cut onwards (right)

4 Discussion

More than half of visual estimates made by experts presented absolute errors above 10 t ha^{-1}. When the acceptable absolute error was defined as up to 10 t ha^{-1}, the estimate in 2015/2016 had an accuracy around 40%, while in 2014/2015, the accuracy was around 50%. Similarly, if we define as acceptable an absolute error of up 20 t ha^{-1}, in 2015/2016 the estimate would have an accuracy around 65%, while in 2014/2015, the estimate would have an accuracy around 80%.

No difference was observed among the estimates for the most frequent sugarcane varieties.

Estimate errors were different for 1st cut compared to 2nd cut onwards. A possible explanation for that is that it is harder to experts to make better estimates for 1st cut; since there is no historical data they tend to make better estimates for subsequent cuts since they are based on previous knowledge.

5 Conclusion

By using REC curves, we can identify the relationships between accuracy and acceptable estimate errors. Estimates for the 1st cut were the worst, and this can be explained by no previous knowledge. By using Decision Tree induction, it might be possible to identify patterns of the errors and understand, for example, why estimates in 2015/2016 were worse than in other harvest seasons.

References

[1] Bernardo Friedrich Theodor Rudorff and Getulio Teixeira Batista. Yield estimation of sugarcane based on Yield estimation of sugarcane based on agrometeorological-spectral models. *Remote Sens. Environ.*, 33:183-192, 1990.

[2] Henrique Coutinho Junqueira Franco, Rafael Otto, Carlos Eduardo Faroni, André César Vitti, Emídio Cantídio Almeida de Oliveira and Paulo Cesar Ocheuze Trivelin. Nitrogen in sugarcane derived from fertilizer under Brazilian field conditions. *Field Crops Research*, 121:29-41, 2011.

[3] Felipe Ferreira Bocca, Luiz Henrique Antunes Rodrigues and Nilson Antonio Modesto Arraes. When do I want to know and why? Different demands on sugarcane yield predictions. *Agr. Syst.* 135:48-56, 2015.

[4] E. Brüggemann, J. Klug, P. Greenfield and H. Dicks. Empirical modelling and prediction of sugarcane yields from field records. In Proc. S. Afr. Sug. Technol. Ass. Page 75, 2001.

[5] Pamela Danese and Matteo Giacomo Maria Kalchschmidt. The role of the forecasting process in improving forecast accuracy and operational performance. *Int. J. Prod. Econ.* 131:204-214.

[6] Jinbo Bi and Kristin P. Bennett. Regression error characteristic curves. In: Tom Fawcett and Nina Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 43-50, Washington (DC), USA, August 21-24, 2003.

SIGRAS App: climate, vegetation and soil information for support systems for decision making in agricultural production through smart devices.

G. Tiscornia[1], A. Gimenez[1], A. Cal[1], J. P. Castaño[1]

[1] Agro-climate and Information System Unit (GRAS). National Institute of Agricultural Research (INIA Uruguay). Ruta 48 KM.10, Canelones, Uruguay. Phone:23677641 int 1758.
gtiscornia@inia.org.uy, agimenez@inia.org.uy, acal@inia.org.uy, jcastano@inia.org.uy

Abstract

One of the most important adaptation measure to Climate Change and Variability is the development of tools and systems to support agricultural activities management and preservation of natural resources. The GRAS Unit of the National Institute of Agricultural Research (INIA Uruguay), elaborate and make available tools and products to contribute to these goals (www.inia.uy/gras). In this regard, "SIGRAS App" was developed, with up-to-date information (estimates of soil water balance variables, NDVI) and historical data (NDVI, soil features, water balance, climate and basic cartography) as well as Tools and Alerts. Information developed with MGAP, INUMET, IRI. Available: Android, iOS and WPhone,

1. Introduction

Among the most important measures to adapt to Climate Change and Variability, is the development of tools and information systems to help farmers, researchers and governments to generate capacities for prevention, adaptation and risk management, and for general management related to agricultural production activities and preservation of natural resources [1] [2]. In this regard and in order to contribute, the Agro-climate and Information System Unit (GRAS) of the National Institute of Agricultural Research (INIA Uruguay) develops, elaborates and makes available to every user and in real time, products, tools and different information [3] [4] [5].

Within the most recently developed products, there is an application for smart devices (cellphones, tablet, etc.) call “SIGRAS App” (Figure 1).

Figure 1. SIGRAS App.

2. Materials and methods

SIGRAS App is mostly based on an on-line web information system “SIGRAS” [6] [7] [8] available since 2013. It access the geographic information database of "SIGRAS" and get the information to be displayed on the mobile device. Through a consultancy, a PHP script was developed that connects to the INIA database and generate a downloadable XML that returns the requested information to the mobile device. The framework PhoneGap (Apache Software Foundation) was used in order to develop the application in HTML5, CSS and JavaScript, and to be able to run it natively on Android and iOS platforms; Offering a unique JavaScript API to access native services. This application needs to access the device's GPS to obtain the current geographical coordinates, for which an HTML5 page was created, making a Javascript call to allow the use of GPS of any device independent of its platform.

This new App provides real time information:

- Estimations of different output variables of a national water balance model with a spatial resolution of 30 km x 30 km, estimated by INIA and methodology describe in [9] (accumulated precipitation, real evapotranspiration, soil available water, well-being water index, potential evapotranspiration, non-retained water)
- Normalized Difference Vegetation Index (NDVI). This index, that uses the visible and near-infrared bands of the electromagnetic spectrum, provides information of the vegetation health as assess chlorophyll activity. We extract the information from the MOD13Q1 MODIS product of LP DAAC, NASA [10]. This product is a composite image, provided every 16 days at 250-meter spatial resolution, with the maximum NDVI value for each pixel in that period.

We also include historical information:

- NDVI statistics based on 11 years information (2000-2011) from 16 days composite images [10]. Average, maximum and minimum values are available for each one of the 23 dates, calculated by INIA.
- Soil characteristics with a spatial resolution of 1:40,000, estimated by a private consultant of INIA, using basic data from de Ministry of Livestock, Agriculture and Fishery of Uruguay (CONEAT index [11], Texture, Depth, Water storage capacity, etc.)
- Water balance statistics based on 25 years (1985-2009) with a spatial resolution of 30 km. x 30 km. Values of the 10, 50 and 90 percentiles are available for each month and each variable (mm - % of soil water, effective precipitation, well-being water index, etc.) estimated by INIA [12].
- Climate statistics based on 30 years (1980-2009) with a spatial resolution of 60 km. x 60 km. Values of the 10, 33, 50, 66 and 90 percentiles are available for each month and each variable: precipitation based on 53 stations and temperature, heliophany, frosty days and relative humidity based on 24 stations, done by INIA and IRI [12]
- Basic cartography (administrative divisions, hydrographic basins, etc), from AGESIC [13].

SIGRAS app also have available a “Tools and Alerts” section were the user can find:

- Six days precipitation forecast and a five days frost forecast elaborated by the Center for weather forecasting and climate studies [14] of Brazil
- Forecast for favorable conditions for fusariosis develop and Deoxynivalenol (DON) mycotoxin in wheat grain, estimated by INIA
- A personalize soil water estimation tool “CuantAgua”, developed by INIA, [15].

Every data is on line all the time, and users can access to the information related to a specific location of interest, choosing from their actual position, the number of the property or just selecting from the map. In the “Marker” item, the user can save the locations they need in order to have them available any time.

1. Results and Discussion

SIGRAS App "packages" integrates data and information from very different sources as remote sensors, historical and real-time databases, model outputs and statistical analyzes. It is one of the first applications, created for smart devices, which allows farmers, researchers and any other users to access to a large amount of information in a very simple and "usable" way in order to contribute to planning and decision-making.

SIGRAS App is free and available to download in on-line Android, iOS and Windows Phone stores for any user.

2. Acknowledgements

This App was developed by the Agro-climate and Information System Unit (GRAS) of the National Institute of Agricultural Research (INIA Uruguay) and includes information elaborated in collaboration with the National Direction of Natural resources of the Ministry of Livestock, Agriculture and Fisheries of Uruguay, the National Meteorology Institute and the International Institute of Climate and society (IRI), Columbia University.

References

[1] Baethgen, W. and Giménez, A. Seasonal Climate Forecast and Satellite Information: improving decisions in the Uruguayan agricultural sector. 2009. In: Climate Sense, World Climate Conference-3, "Climate Predictions & Information for Decision Making", World Meteorological Organization (WMO), Geneva, Switzerland, 31 August-4 September 2009. Published by: Tudor Rose, Leicester, UK.

[2] Giménez, A. y Lanfranco, B. Adaptación al cambio climático y la variabilidad: algunas opciones de respuesta para la producción agrícola en Uruguay. 2012. Revista Mexicana de Ciencias Agrícolas Vol.3 Núm.3 1 de mayo - 30 de junio, 2012 p. 611-620.

[3] GRAS-INIA, 2017. Web page. http://www.inia.uy/gras/

[4] Castaño, JP, Giménez A. y L. Olivera Monitoring natural pastures status in Uruguay using satellite images and a soil water balance model. 2008. International Congress of Pastures and Rangeland, Hohhot, China, 29 de Junio al 5 de Julo de 2008.

[5] Giménez, A., Castaño, J, and Baethgen, W. Development of methodologies and tools for agricultural production risk management. 2010. Fourth International Scientific Conference BALWOIS 2010 (Water Observation and Information System for Decision Support), Ohrid, Republic of Macedonia, 25 to 29 May of 2010.

[6] SIGRAS web, 2017. http://sig.inia.org.uy/sigras/

[7] "Sistema de información geográfica web SIGRAS", A. Cal, A. Giménez y G. Tiscornia, 2013. Revista INIA, Julio de 2013.

[8] G. Tiscornia; A. Cal; A. Gimenez. Sistema de información geográfica web "SIGRAS". Exposición oral. International Conference Climate Services 4. Montevideo, Uruguay. 2014. http://www.climate-services.org/iccs/iccs-4/results.

[9] GRAS-INIA, 2017. National water balance. http://www.inia.uy/gras/Monitoreo-Ambiental/Balance-Hídrico/Balance-hídrico-suelos-Uruguay

[10] LP DAAC, 2017. https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13q1

[11] Grupos CONEAT, 2017. http://www.mgap.gub.uy/unidad-ejecutora/direccion-general-de-recursos-naturales/suelos/coneat/grupos-coneat

[12] Caracterización Agroclimática del Uruguay 1980 – 2009. 2011. Serie técnica 193, INIA, Uruguay.

[13] https://www.agesic.gub.uy/

[14] CPTEC, 2017. http://www.cptec.inpe.br/home/in

[15] CuantAgua: sistema web de estimación personalizada de agua en el suelo. Revista INIA, setiembre de 2015.

Number, maps and facts: Agriculture leads environmental preservation

Evaristo Eduardo de Miranda[1], Carlos Alberto de Carvalho[1], Osvaldo Tadatomo Oshiro[1], Paulo Roberto Rodrigues Martinho[1], Lucíola Alves Magalhães[1], Gustavo Spadotti Amaral Castro[1]

[1]Brazilian Agricultural Research Corporation (EMBRAPA)
Avenida Soldado Passarinho 303, Fazenda Chapadão - Campinas - SP, Brasil
{evaristo.miranda; carlos-alberto.carvalho; osvaldo.oshiro;
paulo.martinho; luciola.magalhaes; gustavo.castro }@embrapa.br

Abstract

This paper presents the role of Brazilian agriculture in the preservation of native vegetation and palustrine and lacustrine ecosystems. After the approval of the New Brazilian Forest Code, farmers are required to register their rural property in the Rural Environmental Registry. By April 30, 2017, more than four million farmers were included, an area of approximately 408 million hectares. The collected data add up to hundreds of millions of polygons, requiring a computational effort and geoprocessing techniques to treat the data and turn it into information. The results indicate that the area reserved to vegetation preservation by farmers represents 20.5% of Brazilian territory, thus demonstrating the importance of the agriculture in environmental preservation.

1 Introduction

Brazilian agriculture has a new and powerful tool: the Rural Environmental Registry System (SiCAR in portuguese). SiCAR is an electronic registry for farms (rural properties) [1], that was created by the Brazilian Forest Service of Ministry of the Environment, as required by the Brazilian Forest Code (Law 12.651/2012) [2]. According to Brazilian Forest Code, in the Amazon biome, for instance, 80% of the farmland should be preserved. Until April 30, 2017, 4,104,247 farms were included in the system, totaling 407,999,690 hectares.

SiCAR includes data about the farm boundary and its internal parts: explored areas, permanent preservation, legal reserve, public infrastructure etc. There are 18 categories of pre-defined land use in SICAR. The boundaries of these categories are defined through Rapid Eye satellite images with 5 meters of spatial resolution. For each category, there is, in general, more than one polygon per farm that delimits it. All these data result in hundreds of millions of polygons, with associated attributes.

The goal of this paper was to analyze SiCAR database in term of the area reserved to preservation and conservation of native vegetation and water bodies in the country by the farmers. The entire SICAR database was processed by experts of the Strategic Territorial Intelligence Group (GITE) of Embrapa. The results reveal the important role of agriculture in environmental preservation and highlight trends about land use and occupation in Brazil.

2 Materials and methods

The data used in this paper were obtained from the website of the Rural Environmental Registration System of the Ministry of the Environment [1]. The download of data from the Ministry of the Environment of each of the 5,570 Brazilian cities resulted in a geocoded archive of more than 40 compacted Gigabyte. The data is provided in 18 different categories, for each farm, but for the purposes of this study, only four categories were considered: permanent preservation areas (PPA), legal reserve (LR), hydrography and native vegetation. In the city of Adamantina (São Paulo state), for example, the file containing the permanent preservation areas (PPAs) of farms has 1,754 polygons. Considering all cities in the state, there are around 800,000 polygons referring only to one of the 18 categories (PPA). Gathering PPAs and legal reserve areas, there are more than one million polygons to be processed to extract the area destined to native vegetation preservation in the 307,000 farms of São Paulo registered in the SICAR.

The following steps were performed to calculate the areas reserved to preservation by agriculture:

- Using the Geographic Information System ESRI ArcGIS, for each of the four categories, the "Repair Geometry" function was applied to correct geometry problems in the categories shapefiles;
- Each shapefile category was clip by the Brazil cities through the function "Intersects";
- After the "Intersects" function, each category was aggregated as a single polygon per city through the "Dissolve" function;
- In order to compose the “preserved vegetation” shapefile, the four shapefiles categories were then unified and the overlays were removed with the function "Union";
- After the unification of the four categories, the “preserved vegetation” shapefile was again re-organized in cities by the "Dissolve" function;
- the amount of areas reserved to vegetation preservation by the city was calculated through the "Calculate Area" function. .

The detailed method is described in the Embrapa project website[1].

3 Results

The Census of Agriculture 2006 [3], last census developed by Brazilian Institute of Geography and Statistics (IBGE), present 5,175,636 farm properties in Brazil. The concept of farm in both contexts, Census and SICAR, is quite similar. By comparing with Census estimate, almost 80% of Brazilian farms were registered in the SICAR (Figure 1) by April 2017. In the Northeast region, more than 1,510,000 units (62%) were not registered. In other Brazilian regions, the number of farm properties is higher than Census. For instance, in the North, the number of registered farmland is 26% higher than those included in the 2006 Census. An increase also occurred in the South (16%), Central West (17%) and Southeast (11%).

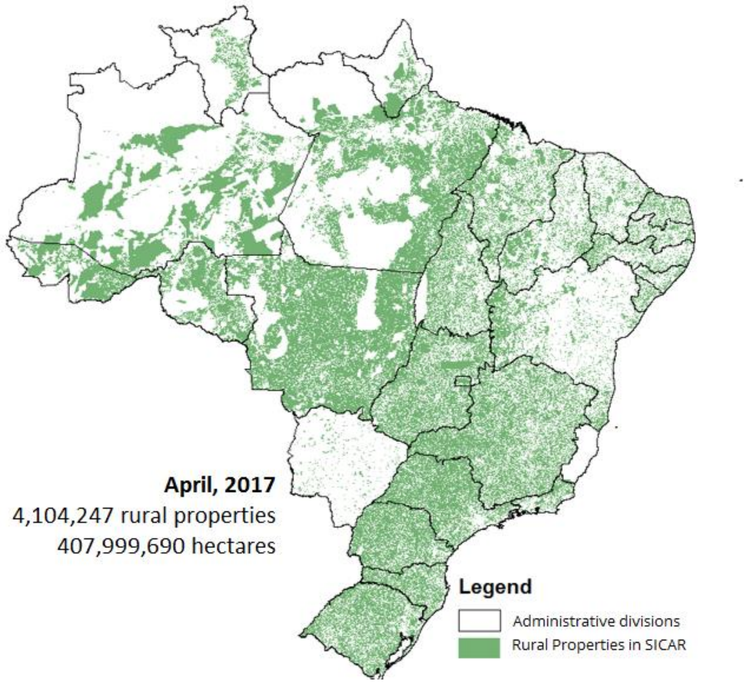


Figure 1: Rural Properties registered in SICAR by April, 2017.

For only two states – Espírito Santo and Mato Grosso do Sul – SICAR data is not available yet. Regarding to the area reserved to vegetation preservation, Southern region preserve 26% of the land, number well above what is required by the Forest Code (20% in that region). The same occurs in all other regions: in Southeastern region the farmers preserve 29%, Central-Western region 49%, North 56%, Northeast 50%.

[1] www.cnpm.embrapa.br/projetos/car/

In São Paulo State (Figure 2), the areas of permanent preservation, legal reserve and surplus vegetation in 309,360 farms, totalize 3,800,000 ha. This area represents more than 15% of the whole state and 22% of the rural area. The area reserved to vegetation preservation in farmlands in São Paulo is higher than all existing conservation units and indigenous lands together (4.5%). Throughout the country, the area reserved to vegetation preservation by farmers represents more than 20% of the national territory, while protected areas represents 13% of the whole country.

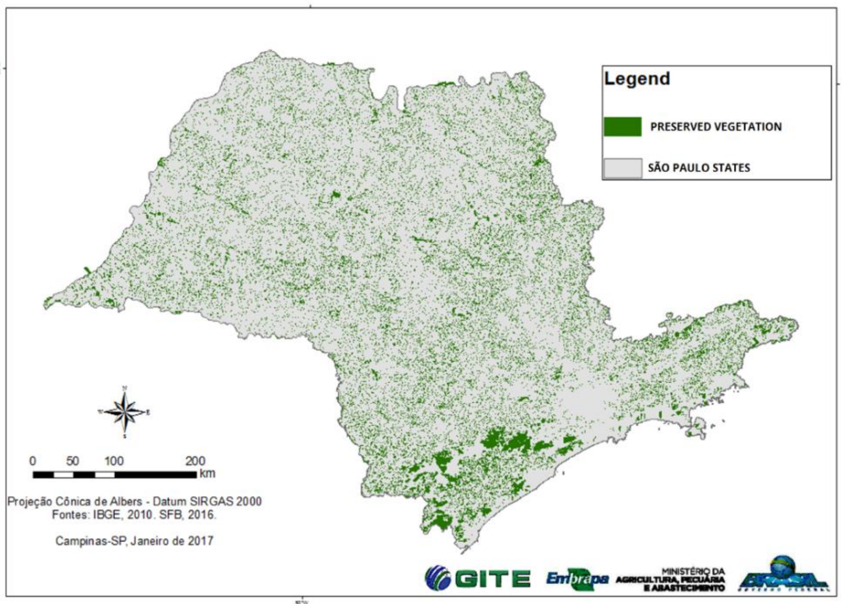


Figure 2: Areas destined to vegetation preservation in São Paulo State.

4 Conclusions

SICAR data shows that no one reserve areas to preserve native vegetation more than farmers in Brazil. The country allocates 66% of its territory to the protection of native vegetation. It would reach almost 75% if pastures located in Pampa, Cerrado, Pantanal, and Caatinga biomes are included. There is no detailed data about the quality of native vegetation [4]. However, the Environmental Regularization Program, to be included in SiCAR, will be adopted to promote the vegetation regeneration process.

The area reserved to vegetation preservation in farmlands represents 20.5% of Brazilian territory, while areas destined to native vegetation protection in conservation units, indigenous lands or unregistered lands represent 13.1%, 13.8%, and 18.9%, respectively.

At Embrapa website, SICAR data can be consulted and analyzed from different spatial perspectives such as cities, states or regions. State maps, aggregated data and summary indicators are also available, demonstrating the key role played by farmers and by agriculture in preserving and conserving native vegetation and waters of the country.

References

[1] Sistema Nacional do Cadastro Ambiental Rural – SICAR. Available at: www.car.gov.br/publico/imoveis/index.

[2] Código Florestal. Lei 12.651 de 25 de maio de 2012. Available at: www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm.

[3] Instituto Brasileiro de Geografia e Estatística (IBGE). Censo Agropecuário 2006. Rio de Janeiro, 2009. Available at: biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf.

[4] Miranda, Evaristo E. Native Vegetation Well Protected by Brazilian Farmers. Available at: https://www.cropscience.bayer.com/en/blogs/corporate-blog/2017/evaristo-eduardo-de-miranda-native-vegetation-well-protected-by-brazilian-farmers.

Gathering spatial data on social vulnerability in Brazil

Luciola Alves Magalhães[1], Marcelo Fernando Fonseca[1], Davi de Oliveira Custódio[1], Paulo Roberto Rodrigues Martinho[1], Jaudete Daltio[1], Carlos Alberto de Carvalho[1], Gustavo Spadotti Amaral Castro[1]

[1]Brazilian Agricultural Research Corporation (EMBRAPA)
Avenida Soldado Passarinho 303, Fazenda Chapadão - Campinas - SP, Brasil
{luciola.magalhaes; marcelo.fonseca; davi.custodio; paulo.martinho; jaudete.daltio; carlos-alberto.carvalho; gustavo.castro}@embrapa.br

Abstract

The Brazilian Ministry of Agricultural and Social Development coordinates several public policies aimed at minimizing social vulnerabilities by means of several programs. These programs' databases are broadcasted to dedicated teams that manage their data using spreadsheets. The need for an integrated view on the territorial coverage of these data motivated a partnership with Embrapa. This paper aims to present the main results of the quantitative, qualitative and cartographic analyses on this data. The resulting geographic data are available in a web platform that may be used in the decision-making process for the expansion or reorganization of governmental social programs.

1 Introduction

The Brazilian Ministry of Agriculture and Social Development (Ministério do Desenvolvimento Social e Agrário, MDSA) is responsible for the guidelines on public policies that handle social inequalities and vulnerabilities. It is equipped with a large database that manages thousands of records and information on several social programs and on the citizens assisted by each of them. Each social program is supervised by dedicated teams which deeply understand the singularities of their actions. However, MDSA currently does not have mechanisms to provide an integrated view on the territorial coverage of its actions and on the correlations between its several social programs.

In order to fulfill this need, MDSA celebrated an agreement with the Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária, Embrapa) for the latter to spatialize these programs' database and correlate them. Embrapa produced quantitative, qualitative and cartographic analyses about the Brazilian cities and the families assisted by the programs.

The aim of this work was to organize and correlate data on MDSA's social programs for strategic territorial analyses, considering two main purposes: to overlap data among different social programs and to enable cartographic analyses by means of a spatial view of the data. Embrapa also detected the need and developed a web platform to host the spatial data and to enable the users to perform queries on it.

Territorially overlapping the data on these programs provides information to support decision making in terms of the distribution of financial resources, the strengthening and/or reorganizing of the assistance to cities and families in need, and improves and optimizes the performance and action range of the governmental agencies.

2 Material and Methods

2.1 Database Organization

The original data were spatialized on the city level. Social programs were analyzed in terms of the number of families assisted and on spatial coverage and were correlated with other programs based on the assisted cities. Cartographic analyses were performed to detect the presence or the absence of social programs in Brazilian cities. Data on the following social programs were used:

- Water-tanks program: Benefits people with limited access to water with the construction of water tanks that capture and store water for use in households, schools or farms;
- Distribution of 'food parcels': Benefits people living under scarcity and vulnerability by means of the distribution of staple food parcels;

- Food security map: Maps the families included in the Single Registry for Social Programs of the Brazilian Federal Government and among which food and nutritional insecurity still persists, considering as main variables (indicators) chronic and acute malnutrition (height-for-age and weight-for-age deficits, respectively) of children under 5 years old along with families assisted by an allowance program;
- Program to foster farming activities: Benefits small farmers by fostering their productive inclusion through financial resources for investments;
- Food Acquisition Program: Small farmers are benefited as preferential suppliers by public entities that buy the food from family farms;
- Seed Warehouse Program: The program is based on the construction of community seed warehouses in the semiarid region, and aims to benefit family farmers that are part of the Federal Government's Single Registry for Social Programs.

Processing and preparing the data for each social program required an amount of effort in order to adapt, correct, convert and make the data compatible for the spatialization process. Data were filtered to identify gaps, duplications, typos, registration errors and differences in spatial location (conflicting information on city and state, for example).

2.2 Spatialization Process

Based on the revised database, the data about each social program were spatialized using the official Brazilian city borders. Layouts containing different themes and territories (region, state, city) and featuring results in the form of aggregated percentages and accumulated totals were produced for each social program in geographic information systems (GIS).

2.3 Web Interface - WebGIS

The spatial data produced was made available through a web interface, named WebGIS. The technology platform that supports the WebGIS relies on open source tools and libraries. The management of the vector spatial data published on the web interface adopts the database management system PostgreSQL[1] and its extension PostGIS, which adds support for geographic objects and operators. The data stored became available via web protocols through geographic web services, according to OGC (Open Geospatial Consortium) standards, using WMS (Web Mapping Service), WCS (Web Coverage Service) and WFS (Web Feature Service) services. We adopted the GeoServer (geoserver.org) map server to provide this functionality. The WebGIS interface was built using javascript Openlayers, ExtJS, Geoext and GXP libraries[2], which are capable of consuming geographic web services, rendering geographic data over maps and providing a range of visual components for user interaction.

3 Results

For each social program, we produced technical reports and maps on a national and state scale, plus the Federal District. The analysis of the spatial coverage of the social programs shows that 57% of the Brazilian cities that show some degree of food insecurity are registered in at least one social program (3,183 cities show a very high, high or average degree, and there is a total of about 1,3 million families in urban and rural areas).

Food parcels were distributed in 1,507 Brazilian cities (27%) and benefited 276,156 families. The program to foster farming activities assisted 215,844 families in 2,055 cities (37%). The water-tanks program built more than 900 thousand structures – 789,635 (82%) for households, 165,461 (17%) for agricultural and animal use, and 3,519 (less than 1%) in schools at rural areas –, and it is available in 1,446 cities (26%) located mainly in the semiarid region, which is traditionally affected by drought periods.

A number of 33,876 families sold their production in the food acquisition program (44,383 tons of food in 2016). The milk supply chain alone exceeded 33 million reais in marketed value, and around 13,430 families in 403 cities are part of these numbers. The seed warehouse program helped about 12 thousand families in 207 cities at the semiarid. Bean is the crop most planted by small farmers, which

[1] postgresql.org | postgis.net
[2] openlayers.org | docs.sencha.com/extjs | www.geoext.org | github.com/boundlessgeo/gxp

shows its importance for food security.

As an example of the correlation of programs, the overlap analysis of the spatial data on food security and food parcels shows that 122 cities under food insecurity were not assisted by food parcels. The original database and the spatial data analysis results are available in a web interface for use by the MDSA team (Figure 1). As soon as the system is validated by their technical teams, it will be available for public use.

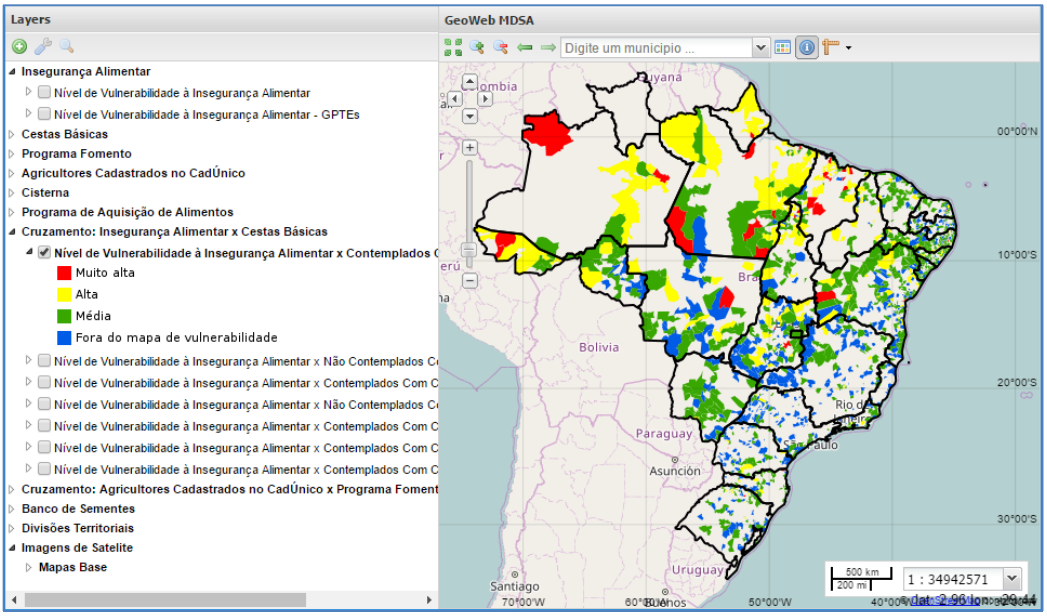


Figure 1 – GeoWeb interface.

4 Conclusions

Brazilian social programs minimize urgent demands, and in many cases promote conditions of self-sufficiency to the affected population. Geoprocessing techniques help map social vulnerabilities and provide spatial knowledge on territory needs and on the coverage of social programs. This spatial perspective helps public managers in the decision-making process in Brazilian cities, based on assertive public policies. The results contribute to identifying priority regions for public actions, by revealing territorial gaps which demand more attention by the public managers. The results provided by the GeoWeb platform allow users with all levels of experience to query and get to know the reality of each Brazilian city assisted by the MDSA.

5 References

[1] MDSA. Ministério do Desenvolvimento Social e Agrário. Portal Brasileiro de Dados Abertos. Disponível em <www.dados.gov.br>

[2] Portal Brasil. Programa leva 54 mil tecnologias sociais a famílias do semiárido, 2016. Disponível: <www.brasil.gov.br/cidadania-e-justica/2016/11>.

[3] Governo facilita acesso a sementes e mudas, 2015. Disponível: <www.mda.gov.br/sitemda/noticias/governo-facilita-acesso-sementes-e-mudas>.

[4] Programa de Aquisição de Alimentos. Manual Operativo, 2014. Disponível: <www.mds.gov.br/webarquivos/publicacao/seguranca_alimentar/manualPAA.pdf>.

IntegraGIS: A GIS system that integrates habitat modelling for vegetable species and the 3PG growth predicting model

A. Márques[1], M. Ortelli[1],
A. Pardo[1], F. Rodríguez[1], D. Strasser[1], P. Hernández[2],
M. León[2], M. Rodríguez[2], D. Valdomir[2]

[1] Universidad Católica del Uruguay
Av. 8 de Octubre 2738, 11600 Montevideo
amarques@dsensetech.com, ortellimarcelo@gmail.com, apardo@ucu.edu.uy,
fran.rp3@gmail.com, distrasser@gmail.com

[2] Abengoa Teyma
Av. Uruguay 1252, 11100 Montevideo
pablohernandez27@gmail.com, maria.leon.r@abengoa.com,
marcelin.rodriguez@gmail.com, dvaldomir@gmail.com

Abstract

IntegraGIS is a system devised to aid in the strategic planning for new international energy forestry plantations. To this end, it integrates environmental data from several sources, including projections into the years 2050 and 2070 that account for the effects of climate change, and applies habitat and growth models on a worldwide extent for hundreds of species. In this document we describe the main aspects of IntegraGIS and its models.

1 Introduction

The purpose of IntegraGIS is to aid in the strategic planning for new energy forestry plantations. In this context it is very important to integrate agronomic data and knowledge into the business process. IntegraGIS aims to provide valuable early-stage analytical information to support the planning process and bridge the communication gap between agronomic engineers and business analysts. The following three main use cases have been identified as central to this effect for an initial version of the system. 1) Given a species, identify suitable areas where it could be grown. 2) Given a site, identify areas with similar environmental conditions. 3) Given a species and a site, predict expected future growth. We describe the approaches employed for these in the following sections.

2 Species habitat modelling

2.1 Land suitability analysis

Land use suitability analysis is the process of determining the level of suitability of a given land area for a certain type of use (agriculture, forestry, etc.). Among the approaches proposed in the literature we have selected a method based on the Analytical Hierarchy Process (AHP) [5] with fuzzy quantifiers because of its widespread use and similarity to preexisting practices at Abengoa. Under this scheme, firstly, suitability is quantified for individual environmental variables (e.g. soil depth, maximum temperature, etc.) using fuzzy functions fitted to model how these variables relate to the abiotic environmental requirements of the species. Then a weighted average of the individual values is computed in order to generate a single overall suitability index. The weights for each environmental variable are determined using AHP, a structured technique for mutiple-criteria decision making, used in this case to formalize expert knowledge of agronomic engineers about the requirements of the species to analyze. For IntegraGIS, an AHP-based model was developed for different varieties of Eucalyptus. For this model, fuzzy quantifiers were fitted to environmental requirement data retrieved from USDA and FAO databases.

2.2 Potential distribution modelling

In potential distribution modelling, location data of sites where the target species has been identified as being present is used with environmental data for those locations to construct a model of the abiotic environmental requirements for said species. This model can then be evaluated on other geographical regions to estimate the degree to which the species' environmental requirements are met. Numerous approaches have been proposed for model construction under this general framework, some specifically designed for this domain, such as Bioclim [4] or GARP [6]. More recently, other general-purpose machine learning algorithms have been applied as well, such as support vector machines (SVM) or random forests (RF). OpenModeller is a cross-platform environment to carry out ecological niche modelling experiments [1]. It includes more than 10 modelling algorithms, including Bioclim, GARP, SVM and RF. We have used OpenModeller for experiments and to build the potential distribution models included in IntegraGIS.

2.3 Experiments

Five species of particular interest for our application were selected to compare these algorithms: Eucalyptus grandis, Eucalyptus robusta, Liquidambar styraciflua, Panicum virgatum and Populus deltoides. Besides the AHP-based model, SVM and RF models for these species were trained on occurrence data downloaded from the Global Biodiversity Information Facility (GBIF), an online repository that serves millions of occurrence records, for thousands of species, from multiple institutions around the world. Occurrences were randomly split 60%/40% into training and testing sets respectively. A matching number of randomly generated absence records was added to each set. The environmental data used in model construction and evaluation was retrieved from multiple sources, including WorldClim, the Harmonized Soil World Database and FAO's GeoNetwork. Our experiments on the occurrence and pseudo-absence data showed that, predictably, machine-learning based methods outperformed the AHP-based model that did not take occurrence data into consideration for its formulation. However, the latter can be a good alternative when either there is no occurrence data available or its quality is too poor. Furthermore, although the AHP-based model performed well for the species for which it was initially developed, its performance degraded for species that should have included other criteria in the model, or for which the environmental variables held different relative importance. SVM and RF produced quite similar results in terms of performance metrics, with RF showing a slight overall advantage. The habitat maps produced by the former had higher regularity than those produced by SVM, and they were also chosen as a better representation of the species habitat by agronomic experts in a subjective blind test.

3 Site similarity

Locating sites with similar environmental conditions to reference high-performing areas can be a very valuable feature when looking for prospective locations for new plantations. To this end, an initial, simple but efficient method to compute site similarity on a global scale has been implemented for IntegraGIS. The range spanned by the environmental variables included in the AHP-based model is computed for the reference area, and each cell of the resulting similarity map is colored according to how many of its corresponding variable values fall within the reference range. For a search area specified by the user, IntegraGIS presents a graph indicating the within-range percentage of cells contained in the search area.

4 Growth modelling

Different approaches have been proposed to model forest growth. The three main types of model found in the literature are: empirical, process-based, and hybrid, which combines the previous two. Empirical statistical models fit standardized functions to large bodies of observed growth data from forest stands. Although empirical models are standard tools for forest managers and can reliably predict growth for the sites and under the environmental conditions for which they were developed, they can be very difficult to extrapolate to sites with different environmental conditions or management regimes. Process-based models aim to predict growth by modelling the internal physiological processes that drive growth based on the species' intrinsic characteristics and the environmental conditions to which stands are exposed. Among these models, we found 3PG [3] to be of particular interest, since it has been successfully tested with numerous species in multiple locations around the world. Recently, Hart et al. have developed and made available through an MIT license a variation of the 3PG model adapted for short rotation woody crops for biofuels: the 3PG-AHB model [2]. The AHB-Poplar application includes environmental data for the northwest of the United States, allowing the manual input of data for other locations. IntegraGIS widens the extent of the available data to cover the whole globe, thus simplifying the model's evaluation for other regions. Besides Poplar, we ran experiments parameterizing the model for Eucalyptus grandis, validating the predicted growth against observed data for reference plantations in Uruguay.

5 Conclusions

IntegraGIS integrates environmental data from multiple sources and applies habitat and growth models on a worldwide extent. For habitat modelling, the system offers a parameterized AHP-based model as well as RF models trained for the 256 species initially included in IntegraGIS. An advantage of the AHP-based model is that, unlike RF models, it does not rely on the availability of species occurrence data for model construction. However, the amount of variables that can be considered by the model before the relative comparison process degrades is limited, fitting optimal fuzzy quantifiers for each of the variables can be very complex, and the resulting models are not trivially adaptable to other species. When reliable occurrence data is available, the RF models were found to give the better results for the evaluated species. The extent of the integrated 3PG growth model's base data was expanded from the US northwest to the whole globe and validated for reference Eucalyptus grandis plantations in Uruguay. These models, as well as a simple approach to determine site similarity, have been packaged into a web application that seeks to enhance the strategic analysis of new international plantations on a worldwide extent.

References

[1] Mauro Enrique de Souza Muñoz, Renato De Giovanni, Marinez Ferreira de Siqueira, Tim Sutton, Peter Brewer, Ricardo Scachetti Pereira, Dora Ann Lange Canhos, and Vanderlei Perez Canhos. openmodeller: a generic approach to species potential distribution modelling. *GeoInformatica*, 15(1):111–135, 2011.

[2] QJ Hart, PW Tittmann, V Bandaru, and BM Jenkins. Modeling poplar growth as a short rotation woody crop for biofuels in the pacific northwest. *Biomass and Bioenergy*, 79:12–27, 2015.

[3] JJ Landsberg and RH Waring. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest ecology and management*, 95(3):209–228, 1997.

[4] Henry A Nix. A biogeographic analysis of australian elapid snakes. *Atlas of elapid snakes of Australia*, 7:4–15, 1986.

[5] Thomas L. Saaty. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, volume 1. McGraw Hill Higher Education, 1980.

[6] David RB Stockwell and Ian R Noble. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and computers in simulation*, 33(5-6):385–390, 1992.

A GIS system to prevent country-wide soil erosion and support sustainable agriculture

Walter Díaz[1], Adrián Márques[1], Alvaro Pardo[1], Javier Preciozzi[1], Santiago Arana[1], Gervasio Piñeiro[1]

Digital Sense Technologies
Michigan 1356 of. 601, 11400 Montevideo
wdiaz@dsensetech.com, amarques@dsensetech.com, apardo@dsensetech.com,
jprecio@dsensetech.com, santiago.arana@gmail.com, gervasiopineiro@gmail.com

Abstract

Uruguay is a pioneer country in soil erosion monitoring, implementing since 2013 a formal process that requires all agricultural activity on an area bigger than 100 ha to be declared to DGRN (Dirección Nacional de Recursos Naturales), the national agency tasked with the conservation of natural resources. Crops must be grown following plans, devised by agricultural engineers, that take into account the expected soil erosion and assure that it remains below its natural regeneration rate. In this document we describe the main aspect of a web-based GIS system built to aid in composing, presenting and controlling these plans.

1 Introduction

Soil is one of the most basic and relevant resources in agriculture. Its preservation from erosion is one of the major concerns of governments from countries all around the globe. The ministry of agriculture (MGAP) of Uruguay implemented in 2013 a national program to prevent soil erosion and push for the sustainability of the different agricultural activities in the country. This requires that all agricultural activity on an area bigger than 100 ha has to be reported and conducted according to plans composed by authorized private-sector agricultural engineers. These plans, named 'Plans for responsible use and management of soils' ('Planes de uso y manejo responsable de suelos' in Spanish), must conform to a series of agronomic criteria, and in particular ensure that soil erosion stays below its regeneration rate. The plans contain information about crop rotation, the management of these crops and data about the soil and terrain. Based on this data, soil erosion is estimated using USLE/RUSLE equations [1, 2] (USLE stands from Universal Soil Loss Equation and RUSLE from Revised Universal Soil Loss Equation). The processes of composing, presenting and controlling plans have been implemented in a web-based GIS system.

2 Developed System

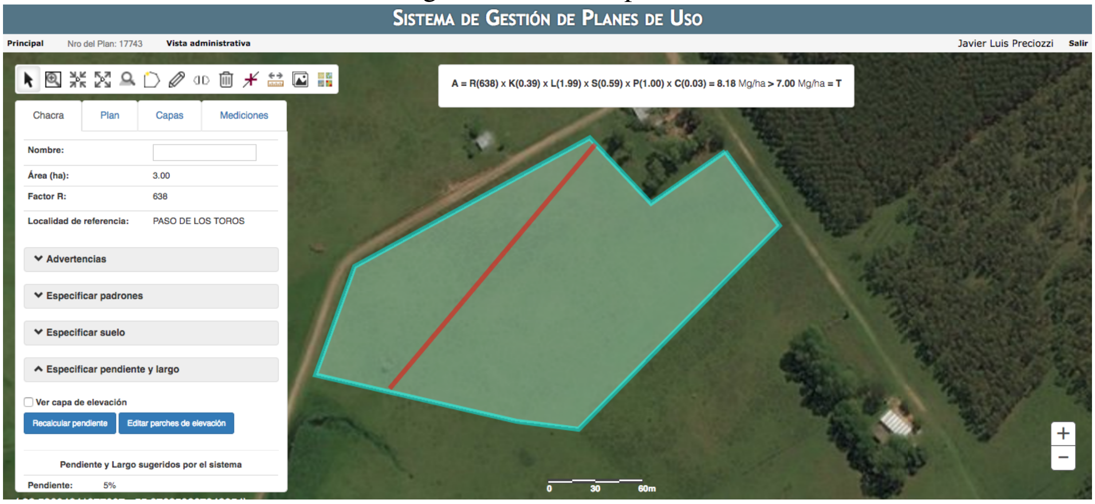
In this work we describe the developed system and its main components. The system has two major user roles, which are the agronomic engineers in charge of presenting the plans (defined as *technical planner*), and the agronomic engineers that work at DGRN on the evaluation and control of the presented plans (defined as *DGRN officers*). The system can also be divided into three main modules:

- **Presentation.** This module is in charge of the presentation of the plans by the *technical planners*.
- **Verification.** In this module, the *DGRN officers* verify the correctness of the presented plans.
- **Control.** Finally, this module is used to control that the presented and verified plans are in fact implemented as declared.

The verification process is heavily based on the USLE/RUSLE equations:

$$A = R \times K \times L \times S \times P \times C < T$$

Figure 1: GIS Component



where R is the rainfall erosivity factor, K is the soil erodibility factor and L and S are topography factors. These four factors depend on the geographical location of the field. P and C are crop management factors that do not depend on the geographical location. The product of these factors, A, is the estimated soil erosion for the field. Presented plans must ensure that this estimated erosion remains below a known bound for the natural regeneration rate for the field's soil type, represented on this equation by T. Because four out of six factors and also T depend on geographical location, the need to implement the system based on a GIS component is clear.

In this work, we will focus our attention only on the GIS components and the proposed solutions to automate and assist the engineers during the presentation of the plans, and the DGRN officers during verification and monitoring.

The system has the following GIS modules: **1)** a component that allows the user to draw the polygons where agricultural activities will take place (see Figure 1), **2)** a module that computes the critical path in terms of slope that it is used to prevent erosion, **3)** several modules to provide georeferenced data such as soil characteristics and climate data, **4)** a module used to monitor the plans against their execution. The latter uses a series of satellite images to extract the NDVI and compare the obtained signature to the one expected based on the declared crops.

In Figure 1 we illustrate the main page of the Presentation module, where modules **1**, **2** and **3** are combined. For each plan, the *technical planner* draws the corresponding polygons for the fields and the system automatically determines the R, K, L, S factors and the tolerance T.

The computation of S and L has interesting details that we explain in what follows. Soil erosion depends on the slope of the terrain and the length of the critical path that starts at one the local maxima of the elevation and ends at the border of the polygon. This path simulates the path that the water will follow from the elevated points towards the border of the polygon. The erosion is measured in terms of the LS factor that captures the effects of the length and slope of the path. The developed system includes a module that automatically computes the critical path and LS factor using a DEM (Digital Elevation Model). For each polygon, the local maxima are extracted and the critical paths automatically computed. The obtained L (length) and S (slope) factors are presented to the user. If the obtained factors indicate that the critical slope is a main driver of erosion in the field, the user can use this information to adjust the plan and compensate for this. These modules greatly reduce the time needed to compute these factors (without them, the user had to compute it offline using her own tools), they standardize the computation and make the process auditable.

Once the plan has been presented it is automatically verified by the system using a set of control rules that trigger alerts. If any of the rules are triggered, a DGRN officer can check the information provided by the technician to evaluate the correctness of the plan. Once the plan has been verified, it is stored for monitoring. The monitoring component of the system uses a temporal series of MODIS images to

Figure 2: Length and Slope determination

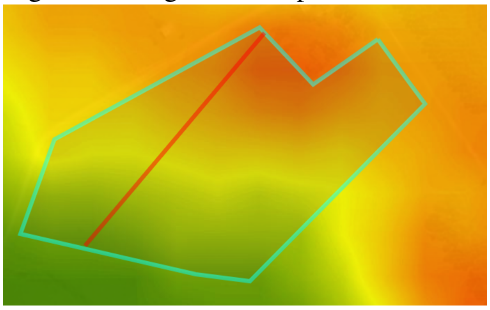
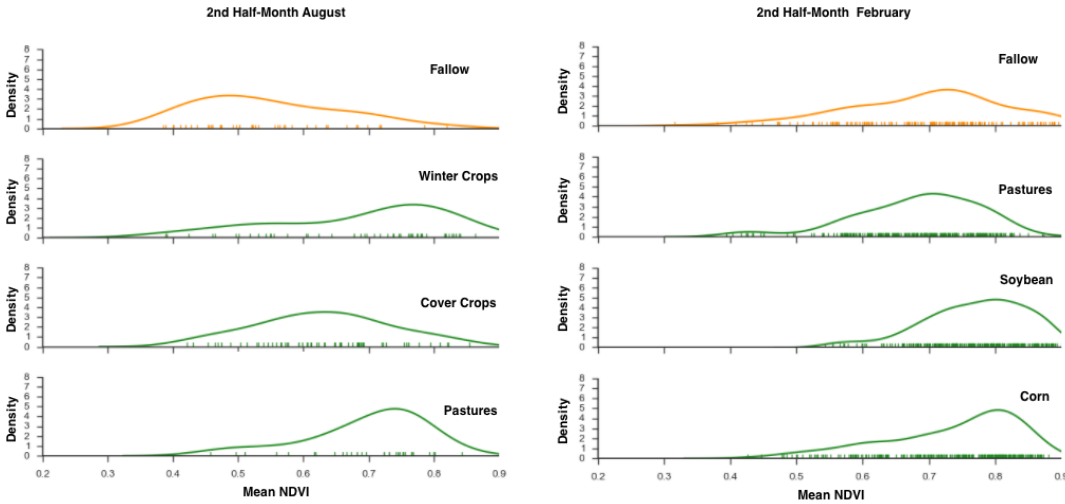


Figure 3: Crop classification based on NDVI



compute the temporal evolution of the NDVI. At key stages in crop evolution, NDVI is evaluated to detect likely deviations from the execution of the declared plan, in particular locating non-covered soil, which is a major driver of erosion. After a detailed analysis of the different crops and the NDVI evolution all along the year, specific dates were selected as the most discriminative, both for summer and winter crops. The NDVI profiles of both of them are presented on Figure 3.

3 Conclusions

Uruguay is a pioneer in terms of public policy to protect soil erosion and pursue sustainable agricultural practices. With over 15,000 declared plans spanning more than 1.5 million productive ha, the developed system is of fundamental importance to enable the composition of accurate plans, to assure data quality and to enable the control and analysis of plans on a country-wide scale by formalizing and automating key processes, such as plan declaration and monitoring via satellite image analysis.

References

[1] A. Durán and F. García Préchac. *Suelos del Uruguay, origen, clasificación, manejo y conservación*, volume II. Hemisferio Sur, 2007.

[2] Walter H Wischmeier, Dwight David Smith, et al. Predicting rainfall erosion losses-a guide to conservation planning. *USDA, Science and Education Administration Agricultiural Handbook N 537, 58p.*, 1978.

Production Planning Model for the assignment of Fermentation Tanks at Wineries

Carlos Monardes[1], Alejandro Mac Cawley[2], Jorge Vera[3], Susan Cholette[4], Sergio Maturana[5]

[1] Pontificia Universidad Católica de Chile
Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile
camonardes@uc.cl

[2] Pontificia Universidad Católica de Chile
Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile
amac@ing.puc.cl

[3] Pontificia Universidad Católica de Chile
Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile
jvera@ing.puc.cl

[4] San Francisco State University
1600 Holloway Avenue, Business Building, San Francisco, California, USA
cholette@sfsu.edu

[5] Pontificia Universidad Católica de Chile
Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile
smaturan@ing.puc.cl

Abstract

We develop a tactical/operational model for planning the assignment of Wineries' Fermentation Tanks. We define a multi-criteria objective function that minimizes a combination of the total number of tank used and total wasted space in the tanks that are used. The model assists the winemaker in the assignment decision of harvested grapes to fermentation tanks, where they will be processed. The decisions are made based on expected yield of the lots where the grapes are grown, harvest dates and fermentation times (days). We present a mixed integer programming model to solve the vat assignment problem, as well as some results obtained from it. Natural extension from this deterministic model are presented explaining the main role of uncertainty and where to incorporate it.

1 Introduction

One of the most worldwide valued industries inside the agricultural sector is the Wine industry. In 2015, Wine industry reached 104.3 million of hectoliters in terms of volume and 28.3 billons EUR in terms of value [6]. In Chile, this industry has experimented a great growth. Last 10 years has doubled its exports [4], reaching 10% of Chilean's exports [2]. Overall, Wine industry has attractive opportunities to Operations Research, because development and application of optimization technics are still growing up [1, 6]. One of the main parts of winemaking, fermentation process, is described below.

1.1 The Fermentation Process

The problem faced by the winemaker before and during the harvest season, is the assignment problem of the harvesting lots to the fermentation vats, to keep the iniquity of the lots and at the same time have an efficient usage of the tanks. He also needs to deal with the uncertainty produced by the climatic factor, which impacts the volume and the dates of harvest, as well as the duration of the harvest. Harvested grapes are the raw material of the fermentation process, which ones have very short duration, so they must be processed as soon as possible. Similarly, if the grapes are harvested before or after their optimal time, the quality of them will probably be affected [3]. The oenologist has the flexibility to advance the harvesting process, to allow more time for the collection and processing of the grapes, with the cost of probably having a lower production and quality. Eventually, it also has the possibility of subcontracting external fermentation tanks to process all the harvested grapes and do not lose part of them.

At the time of planning, the winemaker has a limited number of fermentation tanks, which do not necessarily have the same capacity, but they can be used more than once during the harvest. Depending on the grape lot's volume, it may be necessary to divide this lot and assign it to more than one fermentation tank. In addition, different types of grapes (varietal) cannot be processed in the same fermentation tank, and have different characteristics such as fermentation time or tank's filling space. The fermentation time of the processed grapes in the tanks is a third factor of uncertainty, since different conditions (e.g. ambient temperature, residual sugar in the vats, type of grape, etc.) affect its duration. Next section describes the model.

2 Production Planning Model

2.1 Sets

- $t \in T$: Tanks;
- $l \in L$: Lots of harvested grapes.

2.2 Data Parameters

While some wineries may express tank capacities in terms of volume, such as thousands of liters, we follow the original winery's convention of using tons of grapes crushed.

- C_t : Capacity (tons) of Tank t;
- $P_t = 1$ if Tank t is Proscribed from use in certain fermentations, 0 otherwise;
- S_t : Setup cost incurred each time Tank t is used;
- A_l : expected Amount (tons) of Lot l;
- H_l : expected Harvest date, measured in days from the start of the season, of Lot l;
- F_l : expected Fermentation length (days) of Lot l;
- $R_l = 1$ if Lot l is intended to become a Red wine, 0 otherwise;
- OW : user-specified Objective Weight, where $0 \le OW \le 1$.

2.3 Variables

- $v_{l,t}$: Volume of Lot l assigned to Tank t;
- $w_{l,t}$: Wasted space in Tank t leftover from being used for Lot l;
- $m_{l,t} = 1$ if Lot l is matched to Tank t, 0 otherwise.

2.4 Presolve Calculations

When lots are listed in chronological order of expected harvest, the |**L**| x |**L**| Conflict Matrix will have an upper triangular form.

$$CM_{l_1,l_2} = \begin{cases} 1 & \text{if } H_{l_1} + F_{l_1} \ge H_{l_2} \mid H_{l_1} \le H_{l_2} \text{ and } l_1 \ne l_2 \\ 0 & \text{otherwise} \end{cases}$$

2.5 Objetive Function

Minimize a combination of total tank usage and wasted space within tanks that are used. The user-defined objective weight and setup cost data will influence the solution.

$$\min \; OW \sum_t S_t \sum_l m_{l,t} + (1 - OW) \sum_t \sum_l w_{l,t} \tag{1}$$

2.6 Constraints

All of a lot's tonnage must fit in the tank(s) assigned.

$$\sum_t C_t \cdot m_{l,t} \ge A_l, \forall l \in L \tag{2}$$

As the model tracks both volume used and volume wasted within a tank and may penalize the latter, the next two equality constraints enforce that all tank space is properly accounted for.

$$\sum_t v_{l,t} = A_l, \forall l \in L \tag{3}$$

$$C_t \cdot m_{l,t} = v_{l,t} + w_{l,t}, \ \forall l \in L, \ \forall t \in T \tag{4}$$

This class of constraints prevents a tank from being assigned to two lots that would have a temporal conflict.

$$m_{l_1,t} + m_{l_2,t} \leq 1, \ \forall t \in T, \ \forall l_1, l_2 \in L \mid CM_{l_1,l_2} = 1 \tag{5}$$

The next class of constraints prevents inappropriate lot assignments to tanks with certain proscriptions.

$$\sum_l R_l \cdot m_{l,t} = 0, \ \forall t \in T \mid P_t = 1 \tag{6}$$

Lastly, non-negativity of the continuous variables must be enforced.

$$v_{l,t} \geq 0, \ \forall l \in L, \ \forall t \in T \tag{7}$$

$$w_{l,t} \geq 0, \ \forall l \in L, \ \forall t \in T \tag{8}$$

3 Implementation and Results

The model was implemented in Gurobi 7.5.1 for 64-bit Windows, using Python 3.6 language. The runs were on Intel Core i7-7500U Processor, obtaining solutions in 0.11 seconds (average) for instances with 5 tanks and 6 lots. The model is able to give “good” solutions, by providing an adequate assignment of lots to fermenting tanks in an adequate timeframe.

4 Extension of the model to Stochasticity

The deterministic model faces potential infeasibility of solution when time parameters (harvest and/or fermentation) vary. The winemaker needs robust solutions that keep consistency on matchings, without quality losses. For those reasons, this fermentation process should consider flexibility on vats capacity in order to respond to stochasticity. Since winery works different types of grapes for different final products, it looks natural to lead with variation levels on different type grapes.

The next step to incorporate stochasticity, is to define the robust model, how to measure robustness and associate budget of uncertainty. After that, create three variants of the model (one for each source of uncertainty) and to study the robustness obtained in each one. Finally, incorporate two or three sources of uncertainty (the most relevant ones) and to get final conclusions about the trade-off between robustness and gap from optimality.

5 Conclusions

Our work benefits wineries by providing a tool that will allow the winemaker to quickly determine an optimal schedule and use new information to quickly revise the schedule, reassigning lots that can still be changed. As nature of problem is stochastic, next step is to incorporate uncertainties in the model. That uncertainties arise from climate and nature of grapes fermentation. Climate generates variation in harvest dates and amount (volume) of grape lot, meanwhile nature of grapes generates variation in fermentation times inside tanks.

References

[1] Omar Ahumada and J. Rene Villalobos. A tactical model for planning the production and distribution of fresh produce. *Annals of Operations Research*, 190(1): 339-358, 2011.

[2] Banco Central de Chile. Indicadores de Comercio Exterior, Primer Trimestre de 2017. Reporte Técnico, Departamento Publicaciones de la Gerencia de Asuntos Institucionales. 2017.

[3] Juan-Carlos Ferrer, Alejandro Mac Cawley, Sergio Maturana, Sergio Toloza and Jorge Vera. An optimization approach for scheduling wine grape harvest operations. *International Journal of Production Economics*, 112(2): 985-999, 2008.

[4] Luis Hennicke. Chile: Wine Annual. Global Agricultural Information Network. 2015.

[5] Christos Kolympiris, Michael Thomsen and Justin Morris. An Optimization Model for Winery Capacity Use. In *Southern Agricultural Economics Association Annual Meetings*, Orlando (FL), USA, February 5-8 2006.

[6] Luigi Moccia. Operational Research in the Wine Supply Chain. *INFOR: Information Systems and Operational Research*, 51(2): 53-63, 2013.

[7] The International Organisation of Vine and Wine. STATE OF THE VITIVINICULTURE WORLD MARKET, APRIL 2016. Technical Report, 2016.

[8] Adri van der Merwe, Jan H. van Vuuren and F.E. van Dyk. Decision support for grape harvesting at a South African winery. *ORiON*, 27(2): 83-100, 2011.

Integrated model of crop rotation planning and delineation of rectangular management zones

Virna Ortiz-Araya[1], Víctor M. Albornoz[2], Rodrigo A. Ortega[3]

[1] Grupo de investigación en Agronegocio, Departamento de Gestión Empresarial, Facultad de Ciencias Empresariales, Universidad del Bío-Bío.
Avda. Andrés Bello N°720, Chillán, Chile.
Correo electrónico: viortiz@ubiobio.cl

[2] Departamento de Industrias, [3] Departamento de Ingeniería Comercial,
Universidad Técnica Federico Santa María.
Avda. Santa María N°6400, Vitacura, Santiago, Chile.
Correo electrónico: victor.albornoz@usm.cl, rodrigo.ortega@usm.cl

Abstract

The farmers have to face two important planning problems. The first one is to generate a partition of an agricultural field into rectangular and homogeneous management zones and then to assign correctly a crop rotation for a specific time horizon. The aim of this work is to incorporate these two problems of agricultural planning in an integrated optimization model, whose objective is to minimize the cost of farming in each chosen management zone. This model is implemented using data from an agricultural land of the Bío-Bío region, using AMPL as algebraic language and CPLEX as resolution tool.

1 Introduction

In agriculture, spatial variability of soil properties, such us pH, organic matter, phosphorus, nitrogen, etc., is a key aspect in yield and quality of crops. The new trends in agriculture provide site-specific farming management methods to respond to within-field variability. Based on technologies like sensors and remote sensing, positioning systems, steering systems, geomapping, and integrated electronic communications, it is possible to divide the field into homogeneous management zones. This zoning permits the application of good and efficient agronomic practices for crop planning in a site-specific manner, for example in fertilisation, applying the right amount of nutrients at each zone, in spraying, using an adequate protection from pests and diseases at each zone, and in irrigation, just applying the right amount of water at each zone, among others.

There are several approaches in the literature for determining site-specific management zones. These approaches consider clustering methods based on soil information, see e.g. (Ortega and Santibañez 2007); based on topographic maps, see e.g. (Carr et al. 1991); based on yield maps, combining data from several seasons, see e.g. (Pedroso et al. 2010); and combining them as in Hornung et al. (2006). The combination of different sources of information can be performed by a cluster procedure using for example principal component analysis (Ortega and Flores 1999) or K-means and Fuzzy K-means methods (Jiang et al. 2011). A major drawback is the resulting fragmentation of the zones because these methods usually generate irregular oval or nonconvex shaped zones which do not facilitate the work and operations of machinery. More recently, Cid-García et al. (2013) formulated a binary integer programming model for determining homogeneous rectangular management zones. The model considers spatial variability of a specific soil property and chooses the best field partition from the complete enumeration of all the potential rectangular management zones that it is possible to include using the different soil sample points within the fields. For solving large instances of the proposed integer program, a column generation algorithm was proposed by Albornoz and Ñanco (2016).

The second problem tackled in this paper is to choose the correct crop planning for the different management zones within the partition of the field. This problem is usually referred as the crop planning problem and there are abundant references in literature that are based on Operations Research methodologies, see Memmah et al. (2015). Among many others, Sarker et al. (1997) proposed a linear programming model considering soil type, alternative crops, crop patterns, input requirement, investment, and output. Ortega Alvarez et al. (2004) proposed a nonlinear model solved by genetic algorithms to identify production plans and water irrigation management strategies.

The aim of this work is to incorporate these two problems of agricultural planning in an integrated optimization model, whose objective is to minimize the cost of farming in each chosen management zone.

2 Integrated crop planning model

The proposed crop planning model is based on Santos et al., (2010) ensuring that demand for a specific time horizon is covered by defining a limit for area that will be used. The model assumes a given set of potential rectangular management zones, created according to a computational procedure discussed in Cid-García et al. (2013), as well as a set of possible rotations. More precisely, the sets, parameters and variables are described as follows:

Sets and indexes:

K : set of potential crop rotation plans, with $k \in K$.
S : Set of sample points in the field, with $s \in S$.
T : set of time periods, with $t \in T$.
I : set of crops, with $i \in I$.
Z : set of potential rectangular management zones, with $z \in Z$.

Parameters:

c_{kz} : cost for cultivating the rotation k in the potential management zone z.
a_{sk} : coefficient that takes the value 1 if simple point s belong to zone z and 0 otherwise.
d_{it} : demand for crop i in the time period t.
r^k_{itz} : estimated yield of crop i harvested in period t and crop rotation plan k for zone z.
σ^2 : variance of the soil for the field.
σ_z^2 : variance of the soil property for management zone z.

Decision variables:

$$x_{kz} = \begin{cases} 1, & \text{if crop rotation k is assigned to management zone z en the partition.} \\ 0, & \text{otherwise.} \end{cases}$$

The optimization model is:

$$Min \sum_{z \in Z} \sum_{k \in K} c_{kz} x_{kz} \tag{1}$$

s.t.

$$\textstyle\sum_{z \in Z} \sum_{k \in K} a_{sk} x_{kz} = 1, \quad \forall s \in S \tag{2}$$

$$\textstyle\sum_{z \in Z} \sum_{k \in K} x_{kz} \leq ub \tag{3}$$

$$\sum_{z \in Z} \sum_{k \in K} r^k_{itz} x_{kz} \geq d_{it}, \forall \ i \in I, t \in T \tag{4}$$

$$x_{kz} \in \{0,1\} \ \forall \ z \in Z, k \ \in K \tag{5}$$

$$(1-\propto)\sigma^2 N \geq \sum_{z \in Z} \sum_{k \in K} \left[(n_z - 1)\sigma_z^2 + (1-\alpha)\sigma^2 \right] x_{kz}, \forall \ z \in Z \tag{6}$$

Objective function (1) minimizes the total cost of cultivating the assigned crop rotation at each management zone. Constraint (2) establishes that at each management zone no more than one crop rotation plan can be assigned. Constraint (3) imposes an upper bound to the number of management zones chosen in the field partitioning. Constraint (4) ensures that a specific demand requirement for each crop and constraint (5) defines that all decisions variables must be binary. Finally, constraint (6) requires compliance with a certain level of homogeneity.

3 Resolution methodology and application

This model is implemented using data from an agricultural land of the Bío-Bío region with AMPL as algebraic language and CPLEX as resolution tool (Véliz, 2016). The solved models assume a planning horizon of 4 years divided into annual periods. Different instances were considered, each one of them with a different number of sample points, using organic matter (OM) level as soil property. This last choice, since OM is a key property affecting crop yield and presents spatial and temporal variability in time. For each instance, there are five possible crop rotation alternatives. We will present results for 10 small and medium size instances, considering between 42 and 220 sample points in a field. For large instances, is necessary to design a column generation based algorithm. The proposed methodology is interesting in its latter application, especially when it is considered in terms of a harvest framework for the respective supply chain.

Acknowledgements

This research was supported by DGIIP (Project PIM-17-2) of the Universidad Técnica Federico Santa María and from la DGIP (Project DIUBB Nº161418 3/R) of the Universidad del Bío-Bío. Authors also wish to acknowledge to the IberoAmercican Program for Science and technology for Development (CYTED 516RT0513).

References

[1] Albornoz, V.M. & Ñanco, L.J. (2016). An empirical design of a column generation algorithm applied to a management zone delineation problem. In *Computational Management Science*, Springer, 201-208.

[2] Carr, P., Carlson, G., Jacobsen, J., Skogley, G. & Nielsenand, E. (1991). Farming soils, not fields: A strategy for increasing fertilizer profitability. *Journal of Production Agriculture*, 93, 11–20.

[3] Cid-García N. M., Albornoz V., Ortega R., Ríos-Solís Y.A. (2013). Rectangular shape management zone delineation using integer linear programming. *Computers and Electronics in Agriculture*, 93, 199 – 208.

[4] Hornung, A., Khosla, R., Reich, R., Inman, D. & Westfall, D.G. (2006). Comparison of site-specific management zones: soil-color-based and yield-based. *Agronomy Journal*, 98(2), 407–415.

[5] Jiang, Q., Fu, Q. & Wang, Z. (2011). Study on delineation of irrigation management zones based on management zone analyst software. In: *Computer and computing technologies in agriculture* IV (pp. 419–427). New York: Springer.

[6] Memmah, M.-M, Lescourret, F., Yao, X. & Lavigne, C. (2015). Metaheuristics for agricultural land use optimization. A review. *Agronomy for Sustainable Development*, 35(3), 975-998.

[7] Ortega R., Santibañez O.A. (2007). Determination of management zones in corn (Zea mays L.) based on soil fertility. *Computers and Electronics in Agriculture*, 58, 49-59.

[8] Ortega Alvarez, J.F., de Juan Valero, J.A., Tarjuelo Martín-Benito, J.M., López Mata, E. (2004). MOPECO: an economic optimization model for irrigation water management. *Irrigation Science*, 23, 61–75.

[9] Ortega, R. & Flores, L. (1999). Introducción al manejo sitio-específico. In: Ortega R., Flores, L. (eds) Agricultura de Precisión, Ministerio de Agricultura, Instituto de Investigaciones Agropecuarias. Centro Regional de Investigación Quilamapu (pp.13–46).

[10] Pedroso, M., Taylor, J., Tisseyre, B., Charnomordie, B., Guillaume, S. (2010). A segmentation algorithm for the delineation of agricultural management zones. *Computers and Electronics in Agriculture*, 70(1), 199–208.

[11] Santos L.M.R., anda Arenales M.N., Costa A.M., Santos R.H.S. (2010). Sustainable vegetable crop supply problem. *European Journal of Operational Research*, 204, 639-647.

[12] Sarker, R.A., Talukdar, S. & Haque, A.F.M. (1997). Determination of optimum crop mix for crop cultivation in Bangladesh. *Applied Mathematical Modelling*, 21(10), 621–632

[13] Véliz A.M.I. (2016). Delineación de zonas rectangulares de manejo y planificación de cultivos bajo condiciones de incertidumbre en las propiedades del suelo. Tesis de magíster en ciencias de la ingeniería industrial. Universidad Técnica Federico Santa María.

Resolution of Mixed-Integer Bilevel Problem in the supply chain in meat industry by an Branch & Bound Algorithm

Camila P. Flores[1], Victor M. Albornoz[1], Sara V. Rodriguez-Sanchez[2], Manuel Jimenez-Lizárraga[3]

[1] Departamento de Industrias, Campus Vitacura, UTFSM, Santiago, Chile
camila.flores@alumnos.usm.cl, victor.albornoz@usm.cl

[2] Facultad de Ing. Mecánica y Eléctrica, UANL, Monterrey, México
sara.rodriguezsn@uanl.edu.mx,

[3] Facultad de Ciencias Físicas y Matemáticas, UANL, Monterrey, México
manalejimenez@yahoo.com

Abstract

This paper presents the resolution of a Mixed-Integer Bilevel problem by the application of a Branch & Bound algorithm. The problem to be solved corresponds to a model formulated using Stackelberg's Game theory with application to the supply chain of meat industry. This model focuses on the interaction of two echelon in that chain, proposing the meat packing plant as a leader, and the wholesale store as the follower, seeking to maximize the leaders profits and the level of service of the follower.

1 Introduction

The supply chain of the meat industry involves the interaction of different stackeholders and chain agentes, at different levels. The meat packing is in charge of processing the raw material from carcases to meat products, seeking to maximize profits through the sale of these products, while the wholesalers are in charge of the distribution and marketing of that products. The wholesaler purchases the meat from the producer to meat consumer demand, seeking to maximize its level of service.

The interactions between the two agents in the supply chain are represented by a game theory, with a Stackelberg model, where one agent acts as a leader and the other as a follower [6].

The problem arises when the optimal quantity produced and sold by the meat packing plant is less than the demand required by the wholesaler, and because of this, the wholesaler charges a penalization to the processing company.

The meat packing company,is the first announce a decision make and according to these the follower reacts. A better coordination and integration of activities and the exchange of information among the supply chain agents may reduce the cost given by the penalization imposed by the wholesale.

In [4] it is pointed out that the Stackelberg model can be seen as a Bilevel optimization problem, due to its hierarchical nature. The proposed problem in [1], is presented as a Bilevel optimization problem with integer variables, due the presence of cutting patterns. However, the model was solved by relaxing the integrality of the variables, to reduce the resolution complexity. In this paper the problem will be solved considering the discrete nature of the corresponding decisions variables.

2 Methodology

The problem can be summarized according to the following model in a matrix form:

$$\min_{x\in X} F(x,y) = c_1 x + d_1 y \tag{1}$$

$$s.t.\ A_1 x \leq b_1 \tag{2}$$

$$\min_{y\in Y} f(y) = d_2 y \tag{3}$$

$$s.t.\ A_2 x + B_2 y \leq b_2 \tag{4}$$

In the model, the variables x corresponds to decisions of the leader, being these the level of sales, the quantity to be produced, inventory, carcases to buy, etc. On the other hand, the variables y corresponds to those of the follower, for instance unmet demand and inventory to consider. $F(x, y)$ represents the objective function, that represents the utilities of the leader , where d_1y represents the penalty that the producer receives due to the unmet demand. On other hand $f(y)$ represents the objective function of the follower, modeling the service level granted. Restrictions correspond, in the case of the leader, sales, production and processing balances, as well as capacity constrains such as inventory and time. In the case of the follower the restrictions considered are demand balance and inventory.

To solve Bilevel optimization problems one of the most used approaches is the reformulation by the conditions of Karush Kunh Tucker, that impouse the model of the follower must be convex. The problem nature impouse that the model presents integers variables only in the leader's decisions, so that the follower's problem, by presenting only linear constrains and continuous variables, turns out to be a convex problem. Because of this the follower problem can be represented by the Karush Kuhn Tucker conditions, resulting in a single level nonlinear problem with integer variables.

$$min\ c_1x + d_1y \tag{5}$$
$$s.t.\ \ A_1x \le b_1 \tag{6}$$
$$uB_2 - v = -d_2 \tag{7}$$
$$u(b_2 - A_2x - B_2y) + vy = 0 \tag{8}$$
$$A_2x + B_2y \le b_2 \tag{9}$$
$$x \ge 0,\ y \ge 0,\ u \ge 0,\ v \ge 0 \tag{10}$$

In this new model, $\boldsymbol{x}$ represents the leader's variables and $\boldsymbol{y}$ the follower's variables, $\boldsymbol{u}$ and $\boldsymbol{v}$ represents the shadow prices of the constraints and the variables of the follower respectively. In addition, the constraints (7) and (8), correspond to the KKT conditions that represent the follower problem (3)-(4). In the reformulation, constraint (8) causes the model to cease to be linear, representing the complementarity between the slack's constraints and its respective shadow price, where it is always true that one of these is zero.

In the literature, several approaches have been shown to solve this type of model, as presented in [5]. In [3] is presented an algorithm whose main idea is to solve the model suppressing the complementarity constraint (8) and solving the linear problem; if the point is in the inducible region, it chooses to be a solution of the model, otherwise a branch and bound scheme is used to examine all combinations of complementary slacks. The latter being the one selected to solve the model in study.

3 Algorithmic Strategy

The following is a summary of the algorithm developed for the resolution of the model (5)-(10), based on the presented in [2]. Which was modeled in the algebraic language AMPL.
Step 0:(Initialization) Specifying initial values for the algorithm parameters.
Step 1:(Iteration k) Attempt to solve the iteration's model, if it is infeasible, go to Step 5; otherwise label the solution (x^k, y^k, u^k).
Step 2: (Fathoming) If $F(x^k, y^k) \le \bar{F}$. Go to Step 5.
Step 3: (Branching) If complementary constraint is satisfied, go to Step 4. In contrary case, incorporating restriction to the model, one branch will correspond to include compliance in equality of the restriction, and the other that the shadow price of said restriction is set to zero.
Step 4:(Updating) $F = F(x^k, y^k)$.
Step 5:(Backtracking) Look for those generated models that have not been solved, choosing the oldest one to be solved in the next iteration, go to Step 1. If there are not models unsolved, go to Step 6.
Step 6:(Termination) If $\bar{F} = \infty$, there is no feasible solution to the model. Otherwise, declare the feasible point asociated with $\bar{F}$ the optimal solution.

The criteria for selecting the branching is stablished by the restriction which presents the greatest absolute value between its slack and shadow price. On the other hand the crawl will be from the resolution of the oldest nodes that have been created, generating a search by level.

4 Results

The model will be implemented using data from a real meat packing company operations, considering the use of 17 different cutting patterns, 5 carcases sections, 15 products and a planning horizon of 10 days. Results in a bilevel linear programming model, for which reformulations are presented, becomes a mixed-integer linear problem with 2257 variables in total, divided between those of the leader, the follower and the dual variables. A total of 3471 restrictions, in this it is necessary to consider that in each iteration of the model includes only a part of these; the smallest model, corresponding to the first iteration, has 1157 constraints.
The first instance corresponds to solving the problem with continuous variables to examine the convergence of the algorithm. From results the convergence of the algorithm is confirmed. The resolution times using the algorithm are greater 28,20 times than using approach in [1].
In the second instance, the results are compared by solving the problem by relaxing the integers variables and considering the nature of the variables. In this case, the resolution times for Mixed-Integer problem is 92,59 times grater than the resolution time of Linear problem, because it faces each iteration in a Branch & bound algorithm. The efficiency of the algorithm can be seen notoriously improved or harmed according to the criteria of branching and tracing.

5 Acknowledgment

This research was supported by DGIIP (PIM-17-2), Universidad Técnica Federico Santa María. The authors are grateful for the Programa IberoAmericano de Ciencia y Tecnología de Ciencia y Tecnología para el Desarrollo (CYTED 516RT0513).

References

[1] Javiera A. Bustos, Sebastián H. Olavarría, Víctor M. Albornoz, Sara V. Rodríguez, and Manuel Jimínez-Lizárraga. A Stackelberg Game Model between Manufacturer and Wholesaler in a Food Supply Chain. In *Proceedings of the 6th International Conference on Operations Research and Enterprise Systems*, pages 409–415. SCITEPRESS - Science and Technology Publications, 2017.

[2] Jonathan Bard. *Practical Bilevel Optimization: algorithms and applications*. 2013.

[3] Jonathan F. Bard and James T. Moore. A Branch and Bound Algorithm for the Bilevel Programming Problem. *SIAM Journal on Scientific and Statistical Computing*, 11(2):281–292, mar 1990.

[4] Stephan Dempe. *Foundations of Bilevel Programming*, volume 61 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Boston, 2002.

[5] Jose Fortuny-Amat and Bruce McCarl. A Representation and Economic Interpretation of a Two-Level Programming Problem. *The Journal of the Operational Research Society*, 32(9):783, sep 1981.

[6] Konstantin Kogan and Charles S. Tapiero. *Supply Chain Games: Operations Management And Risk Valuation*, volume 113 of *International Series In Operations Research & Management Science*. Springer US, Boston, MA, 2007.

A Decisions Support System for Purchasing and Storing Fresh Fruit

Wladimir E. Soto-Silva[1], Marcela C. González-Araya[2], Lluís M. Plà-Aragonés[3]

[1] Departamento de Computación e Informática
Universidad Católica del Maule
Av. San Miguel N°3605, Talca -Chile
wsoto@ucm.cl

[2] Department of Industrial Engineering, Faculty of Engineering, Universidad de Talca
Camino a Los Niches Km 1, Curicó - Chile
mgonzalez@utalca.cl

[2] Universidad de Lleida
Jaume II, 73 25001, Lleida - España
lmpla@matematica.udl.cat

Abstract

A decision support system (DSS) to optimize purchasing operations, cold storage selection and fresh fruit freight trips is proposed. The developed DSS seeks to facilitate decisions in a fresh fruit for processing supply chain. Currently, these activities are carried out independently and are not automatized. For this reason, the DSS aims to offer a tactical planning for a processing season. The DSS was implemented in a real case study of a company of dehydrated apples. The implementation shows that it is possible to achieve improvements of approximately 10% in total costs per season.

1 Introduction

The Chilean exports of dehydrated products have shown a rapid growth in the last decades. This is reflected in the total of agro-industrial products exported, where 36% corresponds to dehydrated products, and dried apple in particular has a 7% participation. In 2010, sales were more than US $ 251 million and 122 thousand tons were exported [2]. The growth of dehydrated product industry is also reflected in the participation within the 201 agro-industries registered in Chile in 2005, where 33% is dehydration plants, 29% is frozen product plants, 28% canning plants, and 9% juice plants [3]. Currently, Chile is the first world exporter of dried apples, representing 33.3 % of global exports [1].

In the dried fruit industry, the purchase of a good fresh product and storage are critical issues within the supply chain, since both activities are related to obtaining a good yield in the transformation of fresh fruit into dried fruit [8, 9]. In this sense, the technology of storage selection is relevant, because it is necessary to maintain the fruit quality at the time it is processed [8]. Currently, in Chile there is no a supporting tool to take an optimal decision for the purchasing process, storage and transport of fresh fruit for the dehydration process, so the decisions are made based on the experience of the related stakeholders [9].

Currently, the necessary information for making decisions is stored in spreadsheets and physical documents (reports, books, etc.). Besides, all available fresh fruit in the market is purchased, independently of its quality. This fact causes that the fresh fruit be stored at any available cold chamber, regardless its quality and so, the type of refrigeration technology required. This situation is worsen because there is a shortage of cold chambers during the harvest season and, therefore, the companies rush to hire any kind of cold storage, because they do not have enough cold storage capacity in their facilities.

In this research, a tactical decision support system (DSS) for the purchase, transport and storage of fresh products is proposed, which aims to minimize the costs related to these activities and to maintain the fruit quality as long as possible. The DSS implements the mathematical model proposed by [9]. These authors proposed a mixed integer linear programming model for planning the supply, storage and transport of fresh fruit during a season.

The DSS is applied to a real Chilean case study, corresponding to a processing plant of dehydrated apples. This plant works with approximately 236 fresh fruit growers, and has an annual fruit demand of about 28 million kilograms. The plant possesses 12 cold storage centers, summing a total of 70 cold chambers. These cold chambers have different refrigeration technologies. During the processing season, the company leases a fleet of approximately 30 trucks for the fruit transportation. The trucks are classified in three categories according to their load capacity.

2 Decisions Support System Design

Figure 1 shows the DSS structure, which begins with a mathematical model connected to a database. Both elements are nested in the DSS whose main component is the interface with the user. In order to represent and find solutions, it is necessary to handle input data for the mathematical model, which are the parameters required for modeling the problem, whose storage is performed in a database. In the same way, the outputs of the model called decision variables are managed through a database. The communication with the user is done through an interface, which is responsible for the decision-maker. For the DSS developed, work is carried out with the storage and production area of a process plant, in order to analyze the best alternatives in the management of the input/output of data from DSS.

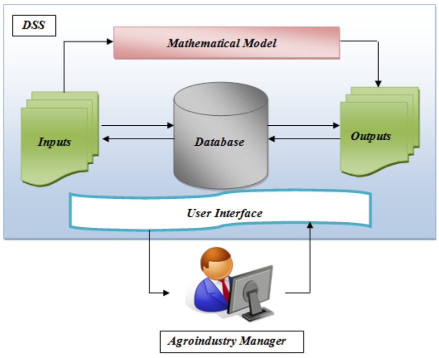


Figure 1: Structure of the DSS (based on [6])

Graphic interface DSS in the programming platform of NetBeans IDE (version 8.0) is created. This tool aims to collect relevant information in order to generate different input parameters for the model. The model is programmed using optimization software ILOG OPL Development Studio IDE (version 6.0.1), that has Cplex software (version 12.6) as an engine of solution.

3 Example of the Results obtained by the DSS for the Case Study

Currently in Chile, the most used tool for storing data is spreadsheet as Excel. For this reason, the data analysis was done in this software. As an example of the different results and reports obtained by the DSS, in Table 1 is presented the purchase report of fresh fruit suggested by the DSS.

.

Fresh Fruit Purchase		Type of quality		
Variety of fruit	Long Quality	Medium Quality	Short Quality	Total Bins
Royal Gala	5,393	3,488	3,119	12,000
Granny Smith	11,139	8,119	14,742	34,000
Fuji	11,761	3,994	6,245	22,000
Brearburn	3,000	-	-	3,000
Pink Lady	2,000	-	-	2,000
Rojas	1,000	-	-	1,000
Total Bins	34,293	15,601	24,106	74,000

Table 1: Results delivered by DSS for fresh fruit purchase

As it can be observed in Table 1, the demand of the processing plant is 74,000 bins of fresh fruit for the season. This demand is satisfied purchasing different varieties and qualities of fresh apple. The DSS also provides reports suggesting the purchase plan according to each fruit producer, the fresh fruit cold storage allocation and the number of truck trips during a season.

4 Conclusions

Usually, among agri-industry processing companies, the size of database can change dramatically depending on the number fruit suppliers and available cold storages. For this reason, the proposed DSS must be flexible enough to be implemented in different kinds of processing plants.

The developed DSS contributes to support complex decisions, which are currently made based on the experience of agribusiness professionals. It was validated according to the opinion of the plant decision makers and its implementation could reduce around 10% of purchase, storage and transportation costs. For the 2010 processing season, this amount corresponds to savings of approximately US$50,000, which were estimated comparing the real operational results with the suggested by the DSS. In addition, the decision makers indicated that the proposed DSS facilitates different activities, such as the order and handling of the information, a greater speed and accuracy for making decisions about purchase, transport and storage of fresh fruit, among others. In this way, a better coordination of these activities would allow to obtain fruit of better quality, improving the conversion rate to dehydrated product and, thus, implying an increase of the product revenue and a reduction of the fruit losses.

Finally, there are many areas in the agricultural supply chain where it is necessary to develop tools to support the decisions [4, 5, 7]. However, despite the fact that agribusiness managers are aware of the lack of decision support tools in the industry, they are reluctant to use DSS because of their little knowledge about the scope and reliability of this kind of tools. Therefore, for the success of a DSS implementation, it is necessary to diagnose different aspects of the agribusiness company, such as current technological infrastructure, financial situation, growth rate expected for medium term and available human capital for the DSS implementation [10]. According to [7], having a good plan and a good identification of the critical factors, it is possible to guarantee the DSS success.

References

[1] Bravo, J. Chile y el mercado mundial de la fruta industrializada. Santiago-Chile: ODEPA, 2010.

[2] Chilealimentos. Diagnóstico Sectorial, Ambiental y Energético. Industria Procesadora de Frutas y Hortalizas. Santiago: Asociación de Empresas de Alimentos de Chile, 2010.

[3] CIREN – Centro de Información de Recursos Naturales. Report N°4: Storage Cold Capacity in Fruit Industry. In Spanish. Download from the link: http://bibliotecadigital.ciren.cl/gsdlexterna/collect/bdirenci/index/assoc/HASH01ba.dir/bolfrut4.pdf, 2005.

[4] Fountas, S., Carli, G., Sørensen, C., Tsiropoulos , Z., Cavalaris,D.,Vatsanidou, B.,Liakos, B., Canavari, M., Wiebensohn, J., Tisserye, B. Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture*, 115:40–50, 2015.

[5] Jain, R., Raju, S. Decision Support System in Agriculture Using Quantitative Analysis. Udaipur: Agrotech Publishing Academy, 2016.

[6] López-Milán, E., Plà-Aragonés, L. A decision support system to manage the supply chain of sugar cane. *Annals Operations Research*, 219:285–297, 2014.

[7] Mir, S., Quadri, S. Decision Support Systems: Concepts, Progress and Issues – A Review. In E. Lichtfouse, editor, *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms*, pages 373-399, Sustainable Agriculture Reviews book series (SARV, volume 2), Springer Science Business Media B.V., 2009.

[8] Oliva, M. Optimization Model for Agribusiness' Apple Supply Chain. Programa de Magíster en Gestión de Operaciones, Facultad de Ingeniería, Universidad de Talca. In Spanish, 2011.

[9] Soto-Silva, W., González–Araya M, Oliva-Fernández, M., Pla-Aragones L. Optimizing Fresh Food Logistics for Processing: Application for a Large Chilean Apple Supply Chain. *Computers and Electronics in Agriculture*, 136:42-57, 2017.

[10] Turban, E. Decision support systems and intelligent systems, 6th edition. Prentice Hall, New Jersey, 2001.

Research Directions in Technology Development to Support Real-Time Decisions of Fresh Produce Logistics

J. René Villalobos[1], Wladimir E. Soto-Silva[2], Marcela C. González-Araya[3], Rosa González–Ramirez[4]

[1] Arizona State University
P.O. Box 875906, Tempe, AZ 85287, Tempe - United States
rene.villalobos@asu.edu

[2] Universidad Católica del Maule
Av. San Miguel N°3605, Talca -Chile
wsoto@ucm.cl

[3] Department of Industrial Engineering, Faculty of Engineering, Universidad de Talca
Camino a Los Niches Km 1, Curicó - Chile
mgonzalez@utalca.cl

[4] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes
Mons. Álvaro del Portillo 12.455, Las Condes, Santiago- Chile
rgonzalez@uandes.cl

Abstract

Recent developments in consumption patterns, lowering of trade restrictions, the emergence of low cost/miniature sensors and information technologies related to business analytics are changing the playing field on which most of the agri-food supply chains operate. The intelligent use of sensing and information technologies have the potential to start a new food revolution in which limited resources could be exploited in an optimal way, reducing food waste. In this research, we review different information systems and technologies used to support real time decision making in each echelon of the process of the export supply chain of fresh fruits.

1 Introduction

The production, transportation, storage and distribution of food will be a critical issue facing humanity in the near future. For instance, the United Nations has warned that to keep up with the increase of the world's population, food production should increase by 70% by the year 2050. In the case of fresh fruits and vegetable, this increase could be even greater since the consumption of these products is associated with health benefits. In this sense, the expanded demand of fresh fruits and vegetables is a challenge and an opportunity. One of the main challenges is how to develop an efficient supply chain that can meet global demand and the intrinsic characteristics of fresh fruits and vegetables at the right prices with minimal waste due to the perishability of these products. In order to better planning and coordination of the fresh fruit supply chains, the use of data collected through the supply chain is very relevant. In this sense, there is an opportunity to use sensors and information technologies whose development is emerging and growing very fast in the last years. In this research, we review these challenges and opportunities in order to improve real-time coordination of different stakeholders of the fresh-fruit supply chain.

2 Challenges and opportunities of the fresh-fruit supply chain real-time decisions

There are a great variety of decision support systems (DSS) that may be implemented to improve the fresh fruit export processes, enhanced by the information data delivered by information technology and sensors downstream and upstream of the supply chain. Hence, it is possible to provide value added services such as traceability of the cargo, market intelligence, real-time logistics, monitoring and control and better planning and coordination of the different stakeholders through the entire fresh fruit supply chain. Table 1 summarizes the state-of-the-art of available technologies to support real-time decisions involved in different echelons of fresh fruit supply chain, where the following abbreviations are used: radio frequency identification (RFID), wireless sensor networks (WSN) and global positioning system

(GPS). In this table, we also identify future technology developments.

Fruit Supply Chain Eche-lons	Sensors	DSS	Related Technologies	Future Technology Developments
Agriculture Practices	RFID [19, 1] WSN [11,18]	– Farm irrigation support system [11, 18]. – DSS for farm fertilization [14].	– Estimation of grape ripeness [3]. – Evaluation of planting recy-cling technologies [5]. – Methodology for using WSN sensors in precision agricul-ture. [9].	- Operational harvest tools tailored to each crop - Automation/ergonomics tools. - Price observatory/data bases. - Protocols for sensor connectivity and networking.
Consolida-tion and Transport Logistics	RFID [16]. RFID, WSN and GPS [20]	– Reefers monitoring [16]. – Monitoring in real time quality products [20]. – Web platform for SCM process optimization [7]. – Food traceability system [6]. – Information system for SCM [10].	– Reefers configuration [8]. – Study of RFID sensors in re-frigerated transport [17].	- Integrated optimization models for location and siz-ing of the consolidation facilities (strategic). - Decision models for design of consolidation fleet (how many, what type) (Strategic/tactical). - Optimal dispatching and routing models for the con-solidation fleet under dynamic/stochastic infor-mation (tactical/operational). - Determination of optimal mix of farmers to attain quality constraints. - Determination of proper time constraints for preser-vation of freshness. - Optimization of overall harvest strategy.
Cold Storage	RFID [19]. Integration of RFID and WSN [2]	– Real-time information system [19]. – Online cold chain track-ing platform [21]. – Online platform sup-ported by different kind of sensors [1].	– Use of artificial intelligence, cloud computing, Geospatial Technology, ZigBee for data analysis [19].	- Integrate cold chain, transportation, and harvest through decision models. - Cold chain monitoring tools. - Low cost appropriate sensor technologies. - Sensor networking and fusion techniques.
Shipping	RFID [17, 13,1] RFID in maritime containers [16]	– System based on RFID and WSN to manage quality in SC [15]. – Smart container combi-nation system [16].	– Intelligent container for refrig-erated transport [12]. – Temperature prediction meth-ods in reefers [4].	- Track the status of the container and plan actions in case the container requires an inspection or an addi-tional documentation when receipt at the port/air-port/border point.

Table 1: Available and future developments of technologies to support real-time decisions in the fresh fruit supply chain.

3 Conclusions

Despite many researches that integrates different technologies to support real-time decisions of fresh fruit supply chain are being developed (see Table 1), there is still a great number of challenges that needs to be dealt with, as suggested in the last column of Table 1. Moreover, there are new technologies that can complement the use of sensors and information systems, such as geospatial technology, data mining, cloud computing, Internet of Things (IoT), artificial intelligence, among others. Recently, new devices to collect data, as nanosensors and drones, are beginning to be used in agriculture. However, there are still little literature about applications and ways to take advantage of all of these new technologies to support real-time decisions in the agricultural supply chain.

References

[1] Abad, E., Palacio, F., Nuin, M., González de Zárate, A., Juarros, A., Gómez, J.M. and Marco, S. RFID smart tag for traceability for traceability and cold chain monitoring of foods: demonstration in an intercontinental fresh fish logistic chain. *Journal of Food Engineering*, 93(4): 394–399, 2009.

[2] Aung, M.M., Liang, F.W., Chang, Y.S., Makatsoris, C. and Chang, J. RFID- and WSN-based intel-ligent cold chain management. *International Journal of Manufacturing Research*, 6(2): 91–109, 2011.

[3] Avila. F., Mora, M., Fredes, C. A method to estimate Grape Phenolic Maturity based on seed im-ages. *Computers and Electronics in Agriculture*, 101: 76–83, 2014.

[4] Badia-Melis, R., Ruiz-Garcia, L. Real-Time Tracking and Remote Monitoring in Food Traceabil-ity. In Montserrat Espiñeira and Francisco J. Santaclara, editors, *Advances in Food Traceability Techniques and Technologies*, pages 209-224, Woodhead Publishing Series in Food Science, Tech-nology and Nutrition, Elsevier, 2016.

[5] Banasik, A., Kanellopoulos, A, Claassen, G.D.H., Bloemhof-Ruwaard, J.M., van der Vorst, G.A.J. Closing loops in agricultural supply chain using multi-objective optimization: A case study of an industrial mushroom supply chain. *International Journal of Production Economics*, 183, Part B: 409-420, 2017.

[6] Chen R.-Y. Autonomous tracing system for backward design in food supply chain. *Food Control*, 51: 70-84, 2015.

[7] de la Fuente- Aragón, M.V., Ros-McDonnell L. Designing a web platform for a fruit-and-vegetable collaborative network. *Procedia Technology*, 9: 247-252, 2013.

[8] Defraeye, T., Nicolai, B., Kirkman, W., Moore, S., van Niekerk, S., Verboven, P., Cronjé, P. Integral performance evaluation of the fresh-produce cold chain. A case study for ambient loading of citrus in refrigerated containers. *Postharvest Biology and Technology*, 112: 1-13, 2010.

[9] Díaz, S.E., Pérez, J. C., Mateos, A. C., Marinescu, M. C., Guerra, B. A novel methodology for the monitoring of the agricultural production process based on wireless sensor networks. *Computers and Electronics in Agriculture*, 76(2): 252–265, 2011.

[10] Djatna, T.,Luthfiyanti, R. An analysis and design of responsive supply chain for pineapple multi products SME based on Digital Business Ecosystem (DBE). *Procedia Manufacturing*, 4: 155–162 2015.

[11] Evans, R. and Iversen, 2008. Evans, R., & Iversen, W. Remote sensing and control of an irrigation system using a distributed wireless sensor network. *IEEE Transactions on Instrumentation and Measurement*, 57(7): 1379–1387, 2008.

[12] Haass, R., Dittmer, P., Veigt, M., Lütjen, M. Reducing food losses and carbon emission by using autonomous control - A simulation study of the intelligent container. *International Journal of Production Economics*, 164: 400-408, 2015.

[13] Hayes, J., Crowley, K., Diamond, D. Simultaneous web based real-time temperature monitoring using multiple wireless sensor networks. In *Proceedings of Sensors, 2005 IEEE*, pages 281-284, Irvine, CA, USA, 30 Oct.-3 Nov. 2005. DOI: 10.1109/ICSENS.2005.1597691

[14] He, J.,Wang, J., He,D., Dong, J.,Wang,Y. The design and implementation of an integrated optimal fertilization decision support system. *Mathematical and Computer Modelling*, 54 (3–4): 1167–1174, 2011.

[15] Hsueh, C.-F., Chang, M.-S. A Model for Intelligent Transportation of Perishable Products. *International Journal of Intelligent Transportation Systems Research*, 8(1): 36–41, 2010.

[16] Jedermann, R., Behrens, C., Westphal, D.,Lang, W. Applying autonomous sensor systems in logistics combining sensor networks, RFIDs and software agents. *Sensors and Actuators A: Physical*, 132(1): 370-375, 2006.

[17] Jedermann, R., Ruiz-García, L., Lang, W. Spatial temperature profiling by semi-passive RFID loggers for perishable food transportation. *Computers and Electronics in Agriculture*, 65: 145-154, 2009.

[18] Kim, Y., and Evans, R. Software design for wireless sensor-based site-specific irrigation. *Computers and Electronics in Agriculture*, 66(2): 159–165, 2009.

[19] Lu. S, and Wang, X. Toward an IT-enabled Business Model for Perishable Food Transportation. *Advance Journal of Food Science and Technology* 11(3): 190-194, 2016.

[20] Wang, L., Kwok, S. K., Ip,W. H. A radio frequency identification and sensor based system for the transportation of food. *Journal of Food Engineering*, 101(1): 120-129, 2010.

[21] Zhang, J., Chen, L., Tang, X. Gao, Q. A Remote Monitoring System for Cold Chain Logistics by Means of Social Networks. *The Open Cybernetics & Systemics Journal*, 9: 888-893, 2015.

A new cloud decision support system for tactical planning in a fruit supply chain

Jordi Mateo[1], Wladimir Soto[3], Marcela Gonzalez[2], Adela Pagès-Bernaus[1], Lluís M. Plà[1] and Francesc Solsona[1]

[1] Universitat de Lleida
Campus Cappont, 25001, LLeida, Spain
jmateo@diei.udl.cat, adela.pages@matematica.udl.cat, lmpla@matematica.udl.cat, francesc@diei.udl.cat

[2] University of Talca
Department of Industrial Engineering, Camino a los Niches km 1, Curicó, Chile.
mgonzalez@utalca.cl

[3] Universidad Católica del Maule
Department of Computer Science, Facultad de Ciencias de la Ingeniería, Av. San Miguel 3605, Talca, Chile.
wsoto@ucm.cl

Abstract

Optimization models are being more used to tackle agro-business problems. This way, cloud computing, machine learning, big data, internet of things and optimization are key factors in the innovation of the sector. This paper proposes a cloud architecture to execute and display models related to the agro-business in a fancy an easy way. Extending the advantages of optimization models with the potential of cloud computing. In this paper, the cloud architecture is designed to help decisions makers in the Chilean dried apple industry to make a better planning over the season.

1 Introduction

Smart agriculture is a development that emphasizes the use of information and communication technology in real life agro-business problems. New technology such as the Cloud computing and the Internet of things are expected to leverage this development and increase the available amount of data stored, the amount of processed real-time data, the introduction of robots and drones and also the artificial intelligence. In [1], an inspiring review focused on smart farming and new technology is presented. There is a rapidly growing literature on this field such as [1], [2], or [3], which indicates the novelty and the increasing interest of combining the emerging technology with the agro-business context.

Cloud computing and Software-as-a-service (SAAS) eliminate the requirement for a powerful computer with an environment dedicated to solve deterministic linear or non linear models. The only thing needed is an electronic device connected to the Internet. With only a web browser, decisions makers obtain access to almost unlimited computing power, no matter which device is used (handheld device, desktop computer or laptop).

In [4], the authors present a deterministic linear model that helps decision makers in the Chilean dried apple supply chain. This work uses this model to show the capabilities, strength, and advantages of merging cloud computing with optimization models. The available evidence seems to suggest that the cloud architecture presented in this paper can be adapted to different kind of models that deals with real-life problems in the context of agro-business.

The application presented in this article offers the possibility of launching executions on the cloud and explore the results in a clear and useful dashboard. It also aims to keep this process as simple as possible. Furthermore, the tool proposed can be fed and merged with the potential of the other technologies required by the sector, such as big data or deep learning.

The major contribution of this work is the design and implementation of a new SAAS optimization service that combines the potential of the deterministic model proposed in [4] and the capabilities of cloud computing. Furthermore, the service is implemented in a user-friendly way to be used without a difficult learning curve. Finally, the service presented is a highly valuable tool. Its flexibility allows this cloud architecture to be used with different optimization models related to the agro-business industry.

2 Cloud Architecure

The platform chosen to develop this architecture was OpenNebula, see [5]. The main reason for choosing this was to take advantage of the Stormy server. This is a private cloud deployed with the OpenNebula platform that belongs to the University of Lleida. See [6] for more information. All the parts of the cloud service were deployed on this platform using Centos7 images as the OS. Furthermore, the architecture presented is very easy to adapt and export to a scalable and commercial environment such as Amazon Web Services or Azure.

The architecture is composed of several parts: the client, the database, the virtual machines dedicated to solve the model and the API-REST that allows the communication between the application and the database. Fig. 1 depicts each part inside the platform.

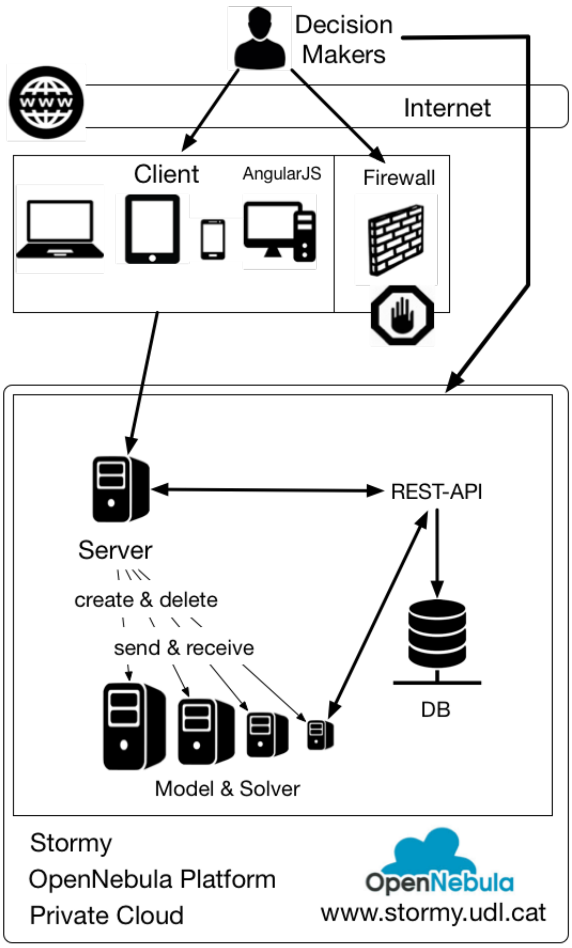


Figure 1: Main parts of the cloud service (SAAS).

3 Modelling of the transportation planning of a fruit logistic centre

Transport planning in a fruit logistic centre is a task with a variable workload depending on the daily demand of fruits, the arrival of new orders and the number of trucks available. The problem modelled here represents the operational planning for day to day. The demand of fruits is defined for the next day and the manager makes the planning with which the activity in the fruit logistic center (FLC) will start the following day. However, new orders may arrive or changes in priorities can be introduced. These unforeseen changes may force to refine or redo the original planning again changing the schedules for truck drivers and suppliers. It is in this context that the model embedded in the cloud system is formulated.

Currently, the model has been tested in a real context, such as is explained in the paper [4]. However, the cloud service proposed in this work has not been tested in a real environment yet.

References

[1] Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. Big Data in Smart Farming – A review. *Agricultural Systems*, 153:69–80, may 2017.

[2] Shuwen Jiang, Tian'en Chen, and Jing Dong. *Application and Implementation of Private Cloud in Agriculture Sensory Data Platform*, pages 60–67. Springer International Publishing, Cham, 2016.

[3] Hyeon Park, Eun-Ju Lee, Dae-Heon Park, Jee-Sook Eun, and Se-Han Kim. Paas offering for the big data analysis of each individual apc. In *Information and Communication Technology Convergence (ICTC), 2016 International Conference on*, pages 30–32. IEEE, 2016.

[4] Wladimir E. Soto-Silva, Marcela C. González-Araya, Marcos A. Oliva-Fernández, and Lluís M. Plà-Aragonés. Optimizing fresh food logistics for processing: Application for a large chilean apple supply chain. *Computers and Electronics in Agriculture*, 136:42 – 57, 2017.

[5] OpenNebula. `http://opennebula.org/`. Online; Accessed: May, 2017.

[6] Stormy. `http://stormy.udl.cat`. Online; Accessed: May, 2017.

A methodology to predict the Normalized Difference Vegetation Index (NDVI) by training a crop growth model with historical data

Andrés Berger[1], Guillermo Ettlin[2], Christopher Quincke[2], Pablo Rodríguez-Bocca[2]

[1] Insituto Nacional de Investigación Agropecuaria
Andes 1365 - piso 12. 11100. Montevideo, Uruguay
aberger@inia.org.uy

[2] Instituto de Computación, Facultad de Ingeniería, Universidad de la República
Julio Herrera y Reissig 565. 11300. Montevideo, Uruguay
prbocca@fing.edu.uy

Abstract

In this paper, we present a methodology to predict the Normalized Difference Vegetation Index (NDVI) by training a crop growth model with historical data. We use a very simple soybean growth model, but the methodology should be able to be extended to other crops and other models. The training process is an optimization problem, that was solved using the spectral projected gradient method. The quality of the prediction was measured by computing the Root-Mean-Square Error (RMSE) between predicted and true values, obtaining an error lower of 9%, that improves the results obtained by simple forecast techniques.

1 Introduction

In the last years, precision agriculture has been promoted in order to increase productivity and to improve the input usage efficiency. One of the most important challenges in precision agriculture is the prediction of crop yields, which allows planning, budgeting, and the precise application of inputs based on future yield. The Normalized Difference Vegetation Index (NDVI) is a measure provided by remote sensors, typically by satellites, drones or ground based sensors, that has a high correlation with crop growth and yield [6].

In this work, we present a methodology to predict NDVI by training a crop growth model with historical data. Crop growth models are computerized representations of crop growth, development and yield, simulated through mathematical equations as functions of soil conditions, weather and management practices [4]. Several crop growth models have been developed, they can be classified into three groups: statistical, mechanistic, and functional models [5]. Our methodology is presented with a very simple functional model developed by G. S. Campbell et al. [2] and adapted by INIA[1] to the soybean crop. The methodology should be able to be extended to other crops and other models. Our model receives as input a set of parameters related to the initial characteristics of the soil, the characteristics of the crop and the meteorological conditions of the place where it is cultivated. The prediction of NDVI by training the model consists of the following steps: first, *calibration:* obtain the values of the parameters relative to the initial characteristics of the soil from the data related to the previous soybean harvests on the fields of study, and from the satellite images and meteorological data; second, *prediction:* use the calibrated model to predict the future harvests (and NDVI series). Our proposal is innovative because we do not use traditional forecast time-series analysis in order to predict the future NDVI series; instead, we calibrate the parameters of a growth model that summarizes the knowledge we have about the real biological process.

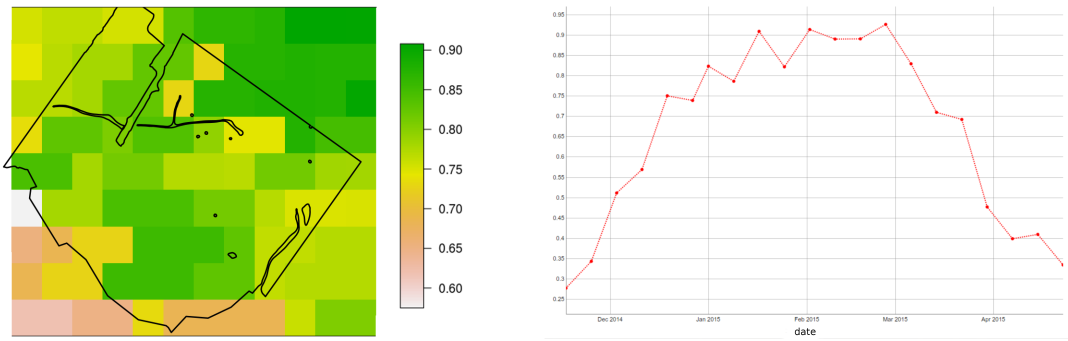
[1]Instituto Nacional de Investigacin Agropecuaria, http://www.inia.uy/, 2017

2 The Normalized Difference Vegetation Index (NDVI) and the Historical Data

The normalized difference vegetation index (NDVI) is a graphical indicator used to assess whether the target being observed contains live green vegetation or not [3]. It is derived from information captured by remote sensors, and it is associated with the fraction of solar radiation absorbed by plants during photosynthesis. It is strongly correlated with green canopy cover, and also in small proportion to the greenness of the vegetation. The formula is in Eq.(1), where NIR and RED stand for the spectral reflectance measurements acquired in the near-infrared regions and red (visible), respectively:

$$NDVI = \frac{NIR - RED}{NIR + RED}. \tag{1}$$

In this work, the NDVI images provided by the MODIS[2] sensors mounted in AQUA and TERRA satellites were used (8 day, cloud free, aggregated product with 250 meter resolution). We computed NDVI temporal series in 37 fields (total: 1945 Has.) geographically distributed in the most important agricultural zones of Uruguay. Considering the NDVI resolution, we have several NDVI time series per field, 2095 time-series in total. For example, in a field named "Churchill", we have one time-series per each square of Figure 1(a), where an example of NDVI time series is in Figure 1(b).



(a) MODIS image of NDVI in the "Churchill" field, where several measure squares (of $250m^2$) are in the field. Date: 25/01/2015.

(b) NDVI time series example of a particular $250m^2$ square in a the 2014/2015 season.

Figure 1: Example of NDVI images and time-series.

For each field we have historical data about 5 soybean seasons, including: planting date, harvest date, soybean variety, the geographical contour of the field, and main meteorological measures per day (max/min temperatures, and precipitations).

3 Prediction Methodology

For each square in each field, the NDVI time-series of a season was predicted, using the historical data and time-series from the other 4 seasons (with a k fold cross validation, in order to reduce overfitting). The data was used to calibrate the parameters of the growth model that was expected to be unchanged between seasons, and that later were used to predict the unknown season. The quality of the prediction was measured by computing the Root-Mean-Square Error (RMSE) between predicted and true values, see Figure 2. The calibration process is an optimization problem, that was solved using the Spectral Projected Gradient method [1], that allowed to include constrained ranges in parameters with agricultural meaning.

[2]Moderate Resolution Imaging Spectroradiometer, https://modis.gsfc.nasa.gov/about/, 2017

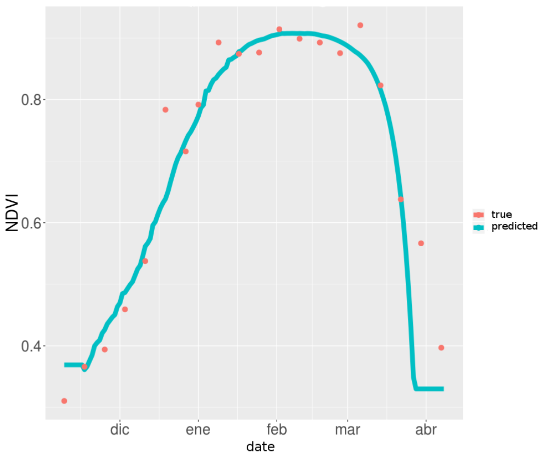


Figure 2: Example of a true and predicted NDVI time-series, using the calibrated growth model and meteorological data.

4 Results and Conclusions

Using this methodology, a RMSE lower of 9% was obtained considering all time-series. The proposed procedure was compared with simple baseline forecast techniques, demonstrating improvements in quality and robustness. For example, robustness was evaluated comparing results under atypical meteorological conditions and considering geographical points outside the fields. Moreover, the obtained parameters about initial characteristics of the soil have agronomic sense, but a validation analysis should be conducted in order to estimate its error.

In order to apply our methodology in future seasons, we need to evaluate it using probabilistic meteorological scenarios. The impact on the quality of this new stochastic data is not estimated yet, and it is an important future work in order to increase the utility of our methodology.

References

[1] Ernesto G Birgin, Jose Mario Martínez, Marcos Raydan, et al. Spectral projected gradient methods: review and perspectives. *J. Stat. Softw*, 60(3):539–559, 2014.

[2] GS Campbell and R Diaz. Simplified soil-water balance models to predict crop transpiration. *Drought research priorities for the dryland tropics. ICRISAT, India*, pages 15–26, 1988.

[3] Robert E Crippen. Calculating the vegetation index faster. *Remote Sensing of Environment*, 34(1):71–73, 1990.

[4] Gerrit Hoogenboom, Jeffrey W White, and Carlos D Messina. From genome to crop: integration through simulation modeling. *Field Crops Research*, 90(1):145–163, 2004.

[5] Kazeem O Rauff and Rasaq Bello. A review of crop growth simulation models as tools for agricultural meteorology. *Agricultural Sciences*, 6(09):1098, 2015.

[6] H Zhang, H Chen, and G Zhou. The model of wheat yield forecast based on modis-ndvi-a case study of xinxiang. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 25–28, 2012.

Design of an early warning and response system for Vegetation fires (SARTiv)

Nicolas Mari[1], Marcelo Scavuzzo[2], Fabiano Morelli[3]

[1] National Commission for Space Activities, Gülich Institute (CONAE-IG)
Spatial Center Teófilo Tabanera, Córdoba, Argentina
National Institute of Agriculture Technology, Cruz del Eje Rural Extension Agency
Eva Perón 451, Cruz del Eje, Cordoba, Argentina
mari.nicolas@inta.gob.ar

[2] National Commission for Space Activities, Gülich Institute (CONAE-IG)
Spatial Center Teófilo Tabanera, Córdoba, Argentina
scabuzzo@conae.gob.ar

[3] National Insitute for Space Research (INPE)
São José dos Campos, San Pablo, Brasil
fabianomorelli@gmail.com

Abstract

Vegetation fires represent a major threat to productive and natural systems; particularly on Natural Protected Areas and forest sector. The information provided by remote sensors offers the possibility of generating very useful tools in the field of prevention, control, and the assessment of fire damages. ***SARTiv*** (Early Warning and Response System for vegetation fires), a Gulich Institute (CONAE/UNC) initiative, seeks to generate information based on the observation and monitoring of the conditions that determine the occurrence of fire cycle. We present its preliminary design and the analysis of users requirements.

1 Introduction

SARTiv, which means "Early Warning and Response System for vegetation fires", emerges as an initiative of the Gulich Institute of CONAE/UNC, to respond to forestry and National NPA´s systems requirements. The project plans to be developed in one inter-agency context with INTA, the National Parks Administration and the National Agriculture, Livestock and Fisheries Ministry, among others.

The system aims to generate information based on the observation and monitoring of the conditions that determine the occurrence of fire cycle (Pre, during and Post). Pre-fire corresponds to the analysis of the conditions of fire risk, such as the accumulation of fuels and their senescence state. The event detection (during), aims to characterize the location and the time of occurrence of the fire, and the fire phase, seeks to quantify the burned surfaces. We present the preliminary design of the system and the analysis of user requirements.

2 Objectives, description and history

The overall objective of *SARTiv* is to generate information for the operational management of vegetation fire based on spatial information. From the point of view of software architecture, the system is designed as a distributed structure of modules, thought to meet the daily demand of satellite products (SP: Systematic Products), and also to generate more specific products defined by end users for specific times and areas on the national territory (ODP: On Demand Products). The everyday processing routine (SP), is aimed to work with satellite data of moderate spatial resolution (e.g.: MODIS, VIIRS data), cartographic sources and numerical models. In parallel, it is expected that the ODP routine provides products at users requirements of high spatial resolution in more detail of the areas and periods of user interests (e.g.: SPOT, Sentinel 2).

In this first stage of design, an analysis of the requirements of the sector corresponding to the management of Protected Natural Areas (PNA´s) and on the forestry sector was done. For both sectors, the specific requirements of spatial information oriented to the activities of prevention, control, and the quantification of damages by fires, were identified. According to such requirements, the basis for the design and the development of products with tailored settings was determined. In this way, the system aims to create products based on the observation and monitoring of the conditions that determine the cycle of the occurrence of fires, whose methodological approach is organized into three main phases: Pre, During and Post fire (PDP). This scheme type is taken from other authors, with the purpose of organizing the processes and the development of identified products to users.

The background in this type of developments has marked a clear path oriented to quick availability of geo spatial products (from satellite origin and other sources), such as maps, alerts, situation reports, and others, via web mapping servers (eg: FIRMS, Queimadas, CONABIO, SEPA). The wide availability of satellite data, along with Geographic Information Systems, and communications technologies, have allowed important advances in the subject, particularly the possibility to access to processed products generated in operational form [6].

3 Architectural design

3.1 *SARTiv* Modules

Danger and fire risk Module (*MoPRI*) must generate various indexes that will identify the state of danger and the Risk of fires at National and Local levels in an integrated manner. At the National level, it intends to work with moderate spatial resolution images (MODIS, 1km) on a daily basis. These images will be used to calculate spectral indices as GVMI (*Global Vegetation Moisture Index*) developed by [2]. This index, combined with human factors, topography and fuels [8], seeks to represent in a dynamic way the conditions of vegetation vulnerability and danger; the latter, associated with the human factor [4]. At the Local level, it intends to replicate the methodology described by means of a parallel processing unit, HRP (*High Resolution Products*), which will operate within the ODP sub-system. This application will respond to the demand for products of high spatial resolution (SPOT, 10m) in predefined areas by the user.

Detection and identification of active fires Module (*MoDIIA*) This module **s**hould identify events of fire through the detection of hot spot groups (cluster), which through a temporary space asociation, take part of the same event of fire [7]. In addition to identifying events, the application must characterize the type of affected vegetation, its location at the provincial and departmental levels, the PNA's involvement; identify the level of risk; quantifying the released energy by the affected coverages (*Fire Radiative Power* - FRP-) and the moment the data was captured [9].

Burn area detection Module (*MoDAQ*) should quantify the surface and the type of coverages affected by fires at National scale and monthly level. The idea is to use MODIS data and follow an adapted methodology from [1]. It is intended that burned area product is distributed on a monthly basis, to a spatial resolution of 500m [3]. Besides, it is intended to have an interface with the HRP processing unit, to meet user demands about areas and specific periods with high resolution, SPOT images 10 m.

Fire behavior and simulation Module (*MoSIMI*) seeks to simulate the behavior of fire from risk conditions of the area of interest chosen by the user [5]. It aims to function as a unit of independent processing within the ODP sub-system. It is intended that the user can perform a remotely simulation analysis, on their area of interest and under custom parameters. The simulation results can be displayed on-line and/or requested for download.

3.2 Products identification

Based on the modules defined and the identified user requirements, the following products are specified:

Raster Maps with 1Km resolution: **Threat Index (*MoPRI_p1*)**, based on ground accessibility characteristics and the presence of human activities; **Topographical vulnerability Index (MoPRI_p2)**, based on topographic features extracted from a Digital Elevation Model representing greater fire warning; **Index of vulnerability of vegetation (*MoPRI_p3*)**, includes the type of fuel and the moisture content state, based on spectral information of satellite origin; **Index of valuation (*MoPRI_p5*)**, indicates the presence of areas with greater value for their fire protection; **Accessibility index (*MoPRI_p6*)**, indicates the chances of access according to the existing communication roads and topographical features; **Integrated index of danger and risk (*MoPRI_p7*)**, represents the conditions of vegetation vulnerability alongside the threats relating to the human factor.

Raster Map with 15km resolution of **Weather vulnerability index (*MoPRI_p4*)**, corresponds to the result of FWI (Fire Weather Index) calculation on the basis of daily variables: Temperature, Relative humidity, Wind speed and Rain (24 hours past), estimated from numerical models of weather forecast (WRF).

Binary Raster Map of **Burned Area (*MoDAQ_p1*)**, includes "burnt/not burnt" classes for the national territory based on images of medium resolution (500 m) of satellite origin.

Vector Maps: Fire events (Cluster Hot Spots) (*MoDIIA_p1*), includes the grouping of fire hot spots (cluster) that together characterize an event of fire; **Simulations of the behavior of fires (*MoSiMI_p1*)**, represents the simulation of burned areas according to the system parameter inputs.

Reports in PDF or another format of **Warnings and summaries of fires at National level (*MoDIIA_p2*)** on the presence of hot spots in the national territory, which shall be made according to the user requirements and distributed via e-mail or via SMS.

4 Concluding remarks

It is supposed that the implementation of this system has a positive impact in areas of decision-making and operating fire management. Moreover, it is expected to obtain users feedback that allows the improvement of the quality of the defined products.

References

[1] Bastarrika, A., Chuvieco, E., Martín, M. P. Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. *Remote Sensing of Environment*, 115, 1003-1012, 2011.

[2] Ceccato, P., Gobron, N., Flasse, S., Pinty, B., Tarantola, S. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sensing of Environment*, Volume 82, Issues 2–3, 188-197, October 2002. ISSN 0034-4257, 10.1016/S0034-4257(02)00037-8.

[3] Chuvieco, E., Opazo, S., Sione, W., Del Valle, H., Anaya, JU., Di Bella, C., Cruz, I., Manzo, L., López, G., Mari, N., *et al*. Global Burned Land Estimation in Latin America using MODIS Composite Data. *Ecological Applications*, 18(1), 64–79, 2008.

[4] Chuvieco, E., Aguado, I., Yebra, M., Nieto, H., Salas, J., Martín, M.P.,Vilar, L., Martínez, J., Martín, S., Ibarra, P., De la Riva, J., Baeza, J., Rodríguez, F., Molina, J.R., Herrera, M.A., Zamora, R. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecological Modelling*, Volume 221, Issue 1, 46-58, 10 January 2010. ISSN 0304-3800, 10.1016/j.ecolmodel.2008.11.017.

[5] Finney, M.A. FARSITE: Fire Area Simulator (Versión 4.1.055) [Software], 2008.

[6] Lentile, LB., Holden ZA., Smith AM S, Falkowski MJ., Hudak AT., Morgan P, Lewis SA., Gessler PE., Benson NC.. Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire*, 15, 319–345, 2006.

[7] Loboda, T.V., Csiszar, I.A. Reconstruction of fire spread within wildland fire events in Northern Eurasia from the MODIS active fire product. *Global and Planetary Change*, Volume 56, Issues 3–4, 258-273, April 2007. ISSN 0921-8181, 10.1016/j.gloplacha.2006.07.015.

[8] Rothermel, R.C. A Mathematical model for Predicting Fire Spread in Wildland Fuels. Research Paper INT-115, Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 1972.

[9] Ressl, R., Lopez, G., Cruz, I., Colditz, R.R., Schmidt, M., Ressl, S., Jiménez, R.. Operational active fire mapping and burnt area identification applicable to Mexican Nature Protection Areas using MODIS and NOAA-AVHRR direct readout data. *Remote Sensing of Environment*, Volume 113, Issue 6, 1113-1126, 15 June 2009. ISSN 0034-4257, 10.1016/j.rse.2008.10.016.

Multi-objective optimization for land use allocation in the basin of Laguna de Rocha

Antonella Barletta[1], Carolina Cabrera[2], Antonio Mauttone[3], Lorena Rodríguez-Gallego[4]

[1,2] Universidad de la República, Facultad de Ciencias
Iguá 4225,11400 Montevideo, Uruguay
abarletta@fcien.edu.uy, carolina@fcien.edu.uy

[3] Universidad de la República, Facultad de Ingeniería
Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay
mauttone@fing.edu.uy

[4] Universidad de la República, Centro Universitario Regional del Este
Ruta nacional Nº9 int. ruta Nº15, Rocha, Uruguay
dunachirca@gmail.com

The use of productive land is subject to pressures and conflicts of interest. In some scenarios, the effect of unplanned land use is associated with the rise of nutrients in the water bodies and the eventual arise of phytoplankton blooms with a dismissal of water quality, which represents a danger for human health [4, 7]. These factors should influence the spatial planning and therefore should be taken in consideration within the process of policy making related to environmental resource management [7]. In this context, the application of mathematical and computational models in land use assignment to collaborate with decision making is an important and increasing field of investigation [3].

We propose a multi-objective optimization model aimed to balance interests related to land productivity and environmental protection. We apply the model to the case of Laguna de Rocha, a coastal lagoon in Uruguay which has been extensively studied since the 1970s. It is considered a protected area by the uruguayan government and it has been recently added as a RAMSAR site [5]. The basin of Laguna de Rocha was transformed into a grid of pixels and its current land use and vegetation coverage have been characterized by a satellite image analysis [6]. For each pixel, the optimization model assigns a land use which can be agriculture, afforestation or cattle raising+conservation. Each pixel has a suitability value for each possible land use. Suitability values were obtained by a multi-attribute model [7]. Moreover, each land use has a value of phosphorus exportation to the lagoon. The optimization model comprises two conflictive objectives: maximization of total land use suitability and minimization of total phosphorus exportation across all pixels. A multi-objective integer linear programming model was proposed, which is solved by using the CPLEX software following two different multi-objective approaches: the weighted-sum and ε-constraint methods [2]. By applying this methodology, we obtain different spatial configurations for land uses in the basin, representing different trade-offs between land productivity and environmental protection.

Figure 1 shows different solutions obtained. When minimizing phosphorus exportation, land uses with smaller exportation values (as conservation) were found in larger coverage areas, whereas when prioritizing the criterion of maximizing suitability, land uses with larger suitability values (like agriculture) were selected by the model. In general terms, we can verify that suitability values and phosphorus exportation values largely determine different spatial configurations of land use. Regarding approaches for solving the multi-objective optimization problem, we found a limitation of the weighted-sum method, which was not able to find solutions within a specific range of trade-off. On the other hand, the ε-constraint method reveals another limitation of the exact modeling proposed, related to the inability of producing spatially compact land use assignments. This can be observed in Figure 2, which shows the results of the optimal solution corresponding to imposing a threshold in the concentration of phosphorus to prevent the growth of potentially harmful phytoplankton [1]. We are currently working on the development of a metaheuristic algorithm able to consider the spatial attributes of the pixels, to generate compact land-use assignments.

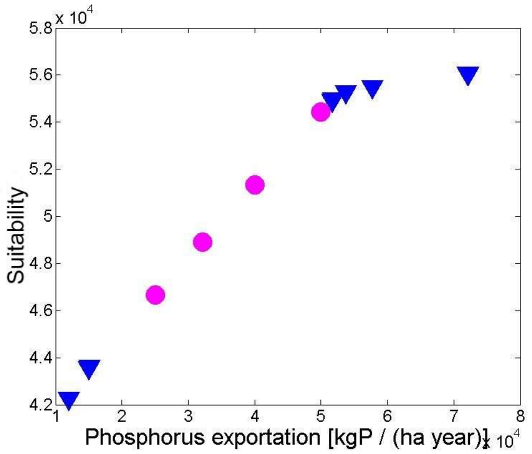


Figure 1: Trade-off solutions obtained by the weighted-sum (blue) and ε-constraint (magenta) approaches

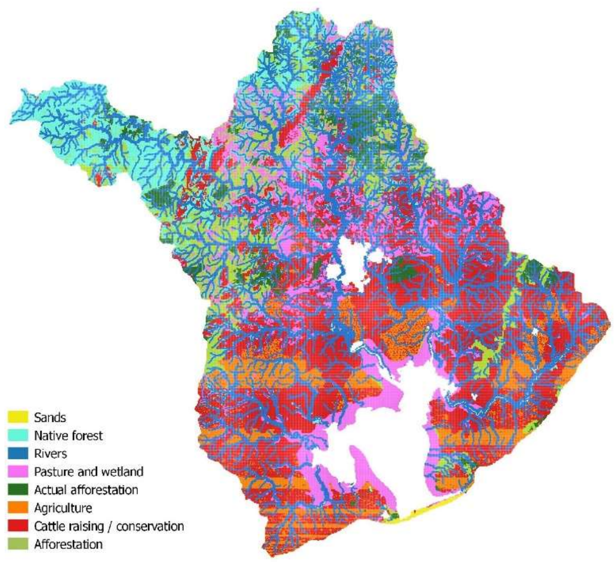


Figure 2: Optimal spatial configuration corresponding to a level of phosphorous exportation which prevents the growth of potentially harmful phytoplankton, 32162.3 kgP / ha year

References

[1] Cabrera, C. Optimización de usos de suelo para prevenir floraciones nocivas de fitoplancton en la Laguna de Rocha, Uruguay. MSc. Thesis, Universidad de la República, PEDECIBA, 2015.

[2] Ehrgott, M. Multicriteria Optimization. Springer, 2005.

[3] de Groot, J. C., Rossing, W. A., Jellema, A., Stobbelaar, D. J., Renting, H., and Van Ittersum, M. K. Exploring multi-scale trade-offs between nature conservation, agricultural profits and landscape quality: a methodology to support discussions on land-use perspectives. *Agriculture, Ecosystems & Environment*, 120(1):58-69, 2007.

[4] Haakonsson, S., Rodríguez-Gallego, L., Somma, A., and Bonilla, S. Temperature and precipitation shape the distribution of harmful cyanobacteria in subtropical lotic and lentic ecosystems. *Science of The Total Environment*, 609:1132-1139, 2017.

[5] Lagos, X., Cabrera, C., and Nogueira L, R.-G. L. Experiencias en el proceso de implementación de un Area Protegida. Apuntes para la Acción II, sistematización de experiencias de extensión universitaria. Extensión Libros, Udelar, Montevideo, Uruguay, 2011.

[6] Nin, M. Mapeo de servicios ecosistémicos en la cuenca de la Laguna de Rocha como un insumo para la planificación territorial. MSc. Thesis, Universidad de la República, 2013.

[7] Rodríguez-Gallego, L., Achkar, M., and Conde, D. Land suitability assessment in the catchment area of four Southwestern Atlantic coastal lagoons: Multicriteria and optimization modeling. *Environmental management*, 50(1):140-152, 2012.

Nonlinear programming techniques and metaheuristics for solving an optimal inventory management model

Fernando Paredes[1] , Javier Pereira[2], Iván Dérpich[3]
Luis Williams[4]

[1]Universidad Diego Portales, Santiago, Chile
fernando.paredes@udp.cl

[2]Tecnológico de Monterrey, Toluca, México
xjavierpereira7@gmail.com

[3]Universidad de Santiago de Chile, Santiago, Chile
ivan.derpich@usach.cl

[4]Universidad San Sebastián, Santiago, Chile
luis.williams@uss.cl

Abstract

This paper compares strategies to solve a specific optimal management model for a large size inventory, determining the appropriate quantities and safety stocks for a periodic inventory management system, and ensuring that the expected lost sales are minimized for a given level of budget and number of annual replenishment orders. Specifically, the results obtained by solving this non-linear programming model based on a fixed-point solving strategy applied to the Karush-Kuhn-Tucker optimality conditions, are compared with those obtained by the application of Tabu Search and Particle Swarm metaheuristics. The computer results obtained so far are quite promising, when comparing the quality of the solutions found, and their respective solving times.

Keywords: Inventory Management, Non-Linear Programming, Metaheuristics.

1 Introduction

Most product distributors commonly face a highly competitive environment; therefore, it is vitally important to manage inventory levels efficiently in order to ensure adequate customer satisfaction and, at the same time, minimize the financial investment required. This problem has been studied in the literature by different authors and has naturally led to the development of several optimal inventory management models; see, for example, [2], [4], [8], [10], among others.

This paper describes a specific optimal inventory management model [5], which determines the optimum order quantities and moments for the periodic management system of a large size inventory, ensuring that the expected lost sales are minimized for a given level of financial investment and annual number of replenishment orders. The difference with other approaches, whose emphasis is mainly on ensuring an adequate level of service, is fundamentally that this paper addresses the commercial problem associated with lost sales, which are not necessarily related to the level of service for a given financial investment. Solving this non-linear programming model based on a fixed-point solving strategy applied to the Karush-Kuhn-Tucker optimality conditions [5], [9], is compared with solving it through the application of Tabu Search and Particle Swarm metaheuristics [6]. The computer results obtained so far are quite promising when comparing the quality of the solutions found, and their respective solving times. A direct extension of this work to the optimal management of resources in the food industry can be done by closely following the work developed in [3] in which optimal management strategies are proposed in the production of Canned Sweet Corn Kernel Industry.

2 Model formulation

The optimal inventory management model consists of minimizing the expected annual lost sales. For this model to represent the financial aspects of the company in an attractive way, the lost sales were evaluated considering the margin of the products and not the sale price. It should be noted that if the preceding consideration were not applied, the total annual sales would be maximized, but the total annual margin might not be at its optimal value due to differences in sales prices, expenses and margins.

An important constraint of the model is based on the annual total inventory investment defined as the average inventory level considering the purchase cost. This constraint represents the typical situation in which a commercial distributor has to implement inventory management policies under a limited budget. Another important constraint of the model is the annual average frequency of product replenishment, that is, the average number of times that the replenishment of the products is ordered annually.

Applying specific results [1], [4], the following optimal inventory management model is obtained:

$$\min \quad z = \sum_{i=1}^{N} \sigma_{L_i} G(k_i) \frac{D_i}{Q_i + \sigma_{L_i} G(k_i)} m_i$$

$$\sum_{i=1}^{N} \left(\frac{Q_i}{2} + \sigma_{L_i} k_i + \sigma_{L_i} G(k_i)\right) c_i \leq I$$

$$\sum_{i=1}^{N} \frac{D_i}{Q_i + \sigma_{L_i} G(k_i)} \leq O$$

$$Q_i \geq 0 \ , \ \forall\, i = 1,2,\ldots\ldots, N$$

$$k_i \geq 0 \ , \ \forall\, i = 1,2,\ldots\ldots, N$$

where:

N = number of different products in stock annually,

$D_i =$ expected demand for product i per year,

$m_i =$ unit margin of product i per year,

Q_i = size of the replenishment order for product i,

$L_i =$ lead time for product i replenishment order,

$\sigma_{L_i} =$ standard deviation of the lead time for product i replenishment order,

$c_i =$ unit cost of product i per year,

d_{L_i} = expected demand during product i replenishment lead time,

$r_i = d_{L_i} + k_i \sigma_{L_i}$, product i reorder point with k_i, the corresponding safety factor,

I = annual investment,

$O =$ annual number of replenishment orders

$G(k_i)$ = product i loss function according to its safety factor, defined by the unit normal distribution:

$$G(k_i) = \int_{k_i}^{\infty} (u - k_i) \frac{e^{\frac{-u^2}{2}}}{\sqrt{2\pi}} du$$

3 Model solving and results

The diverse results obtained by solving this non-linear programming model based on a fixed-point solving strategy applied to the Karush-Kuhn-Tucker optimality conditions [5], [6], are compared with those obtained by the application of Tabu Search and Particle Swarm metaheuristics, [7]. The computer results obtained led to quite promising solutions when comparing their quality and respective solving times.

References

[1] A. Hax, D. Candea. Production and Inventory Management. Prentice-Hall Inc, 1984.

[2] B. E. Flores, D. C. Whybark. Implementing Multiple Criteria ABC Analysis. *Journal of Operations Management*, Vol. 7, 79-84, 1987.

[3] C. Ongkunaruk, Prawat, R. Inventory Management in Canned Sweet Corn Kernel Industry. *Proceedings of the 21th International Conference on Flexible Automation and Intelligente Manufacturing (FAIM), Taiwan,* 2011.

[4] E. Silver, D. Pyke, R. Peterson. Inventory Management and Production Planning and Scheduling, Third edition, John Wiley & Sons, 1998.

[5] F. Paredes, J. Pereira, C. Fuentes. An inventory optimization technique based on minimizing lost sales. Euro-Informs Joint International Meeting, Rome, Italy, 2013.

[6] F. Paredes, J. Pereira, C. Fuentes, B. Crawford, R. Soto. A new algorithmic approach for solving an inventory optimization model, 20th Conference of the International Federation of Operational Research Societies , Barcelona, España, 2014.

[7] G. Cabrera, P.A. Miranda, E. Cabrera, R. Soto, B. Crawford, J. Rubio, F. Paredes. Solving a Novel Inventory Location Model with Stochastic Constraints and (R,s,S) Inventory Control Policy. *Mathematical Problems in Engineering*.Volume 2013, Article ID 670528, 12 pages, 2013.

[8] J. A. G. Krupp. Are ABC Codes an Obsolete Technology?. *APICS – The Performance Advantage,* April, 34-35, 1994.

[9] R. T. Rockafellar, Lagrange Multiplier and Optimality, SIAM Review, Vol. 35, 183-238, 1993.

[10] R. Zhang, W. Hopp, C. Supatgiat. Spreadsheet Implementable Inventory Control for a Distribution Center. *Journal of Heuristics*, Vol. 7, 186-203, 2001.

Optimization in the planning of forest harvest services

Víctor Viana[1], Héctor Cancela[2], Lorena Pradenas[3]

[1] Universidad de la República, Sede Tacuarembó
Ruta 5, km. 386,5 – Tacuarembó, Uruguay
vviana@cci.edu.uy

[2] Facultad de Ingeniería, Universidad de la República
J. Herrera y Reissig 565, Montevideo, Uruguay
cancela@ ng.edu.uy

[3] Departamento de Ingeniería Industrial, Universidad de Concepción
Concepción, Chile
lpradena@udec.cl

Forest planning is a highly complex decision-making problem involving the participation of various factors: ecological, productive and economic systems ([1], [2], [3]). The management of the forest resources involves decision making at strategic, tactical, and operational levels, with both short and long term consequences. One of the components of this decision making chain corresponds to forest harvesting, which is a step of vital importance because of its high impact on production costs. An e cient planning of harvesting operations can signi cantly lower costs associated with logistics and improve the economic performance of companies in the sector.

In Uruguay, almost 75% of the total forestry operations destined to the production of pulp are in charge of contractor companies, which are an important player in the supply chain. This work, which is part of a M.Sc. thesis, aims to optimize the scheduling and routing of the contractors' harvesting equipment, which must be relocated between the places to be harvested in a season.

Currently, most forests in Uruguay are devoted to producing wood for pulp mills. The trees can be harvested after they reach the age of eight years. There are some large pulp companies, which are also owners of most of the tree plantations, and who have agreements with contractor companies to carry out the harvesting. Each contractor company is assigned an annual harvest plan consisting of a series of plantations which it must visit and harvest the areas requested, committing in each case to:

- run that harvest in a certain period of time,
- deliver a given volume of wood in each period.

At each visit to the harvesting areas or blocks, the contractors have to locate the necessary equipment to carry out the cutting and loading operation. The contractors may decide the order in which to visit the plantations; an e cient choice should minimize the transfer costs of the equipment between the harvest sites. Moreover, each harvesting equipment belonging to a contractor starts its journey from a base of operations of that company; and it returns there once the whole harvest is nished.

This work proposes a combinatorial optimization model for this problem, based on the Multi Depot Multiple Traveling Salesman Problem (MmTSP). Numerical experiments have shown that instances of realistic size can be solved by standard mathematical programming software in a reasonable time. This model can then be used as an e cient and practical tool for helping to develop the annual forest harvesting scheduling and equipment allocation for contractor companies, leading to better quality plans and improvement opportunities.

References

[1] Weintraub, A., and Romero, C. (2006). Operations research models and the management of agricultural and forestry resources: a review and comparison. Interfaces, 36(5), 446-457.

[2] Milanesi, G. S., Broz, D., Tohmé, F., Rossit, D. (2014). Strategic analysis of forest investments using real option: The fuzzy pay-o model (FPOM). Fuzzy Economic Review, 19(1), 33.

[3] Broz, D., Durand, G., Rossit, D., Tohmé, F., Frutos, M. (2016). Strategic planning in a forest supply chain: a multigoal and multiproduct approach. Canadian Journal of Forest Research, 47(999), 297-307.

[4] I. Kara and T. Bektas, "Integer linear programming formulations of multiple salesman problems and its variations," Omega, vol. 174, pp. 1449–1458, 2006.