Computers and Electronics in Agriculture 148 (2018) 237–249

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

A new decision-support system for the historical analysis of integrated pest management activities on olive crops based on climatic data

Claudio Zaza[a,*], Sandro Bimonte[b], Nicola Faccilongo[a], Piermichele La Sala[a], Francesco Contò[a], Crescenzio Gallo[c]

[a] Department of Economics, University of Foggia, 1, Largo Papa Giovanni Paolo II, Foggia, Italy
[b] TSCF, Irstea, 9 Av. Blaise Pascal, Aubiere, France
[c] Department of Clinical and Experimental Medicine, University of Foggia, 1, Viale Luigi Pinto, Foggia, Italy

ARTICLE INFO

Keywords:
IPM
OLAP
GSI
Advanced multidimensional model
Olive tree

ABSTRACT

Olive tree is one of the most important crop at global scale. Apulia is the first olive-producing region in Italy, with a huge amount of farms that generate Integrated Pest Management (IPM) data. IPM requires the simultaneous use of different crop protection techniques to control pests through an ecological and economic approach. The crop protection strategies are correlated to the climate condition considering the very important relation among climate, crops and pests. Therefore, in this work is presented a new advanced On-Line Analytical Processing (OLAP) model integrating the Growing Season Index (GSI), a phenology model, to compare indirectly the farms by a climatic point of view. The proposed system allows analysing IPM data of different farms having the same phenological conditions over a year to understand some best practices and to highlight and explain different practices adopted by farms working in different climatic conditions.

1. Introduction

Olive tree (*Olea europea* L.) is one of the most important crop at global scale. The cultivated surface at world level is equal to 10.27 Mha with a production of 15.4 Mt (FAOSTAT, 2014). The leading three countries for Olive crop invested area are Spain (2.52 Mha), Tunisia (1.59 Mha) and Italy (1.15 Mha), while at production level there are Spain (4.56 Mt), Italy (1.96 Mt) and Greece (1.78 Mt) (FAOSTAT, 2014). According to the Italian Institute of Statistics (ISTAT, 2016), Apulia is the first olive-producing region in Italy. In particular, in 2016 the agricultural surface invested in Apulia for olive production was equal to 0.379 Mha, thus representing 33% of the corresponding national rate (1.14 Mha). The obtained production was 0.99 Mt, corresponding to 35% of the national amount (2.81 Mt) (ISTAT, 2016).

Therefore, in the Apulia region, a large number (227245) of farms works on the Olive crop (ISTAT, 2010), which generates a huge amount of *Integrated Pest Management* (IPM) data. Actually, these data are quite unexploited. The IPM approach consists of the application of different crop's protection methods in order to maintain the pest populations below an economic damage threshold (Chandler et al., 2011). Therefore, the IPM crop protection strategies are composed of different techniques to control insects, pathogens, weeds and vertebrates, with an ecological and economic approach (Prokopy, 2003). With the Framework Directive 2009/128/EC (European Parliament, 2009) on Sustainable Use of Pesticides, the European Union (EU) has highlighted that the application of the IPM approach could be the answer to decrease the reliance on the use of conventional pesticides. Therefore, since the 1st January 2014 in the European member countries is mandatory for all the professional users of pesticides to implement the IPM strategies. The Framework Directive 2009/128/EC requires that each Member States must release the crop technical specifications, combining the eight IPM principles (Barzman et al., 2015).

Nowadays with the democratization of information technologies, agriculture enquiries more and more the usage of these technologies in all its practices. Therefore, *Farm Management Information Systems* (FMISs) have been developed. FMISs are systems able to carry out functionalities for field operations management, best practice tools, finance, inventory, traceability, reporting, site-specific tools, sales, machinery management, human resource management, and quality assurance (Fountas et al., 2015). In many commercial FMISs, the current functions related to best practices, such as adherence to organic standards or IPM requirements, are still in their infancy (Fountas et al., 2015).

Taking into account the *environmental* (i.e. *climatic) factors* is a

* Corresponding author.
E-mail addresses: claudio.zaza@unifg.it (C. Zaza), sandro.bimonte@irstea.fr (S. Bimonte), nicola.faccilongo@unifg.it (N. Faccilongo), piermichele.lasala@unifg.it (P. La Sala), francesco.conto@unifg.it (F. Contò), crescenzio.gallo@unifg.it (C. Gallo).

https://doi.org/10.1016/j.compag.2018.03.015
Received 8 February 2018; Received in revised form 10 March 2018; Accepted 16 March 2018
Available online 28 March 2018
0168-1699/

mandatory task to create a real useful instrument to help farmers to analyze the management choices. In fact, the environmental variables, such as air temperature, relative humidity, solar radiation, rain, etc., interacting constantly with the plants and the pests, as reported by many authors in literature. In particular, radiation, temperature, and water have a great impact on the growth of plants (Hay and Porter, 2006).

As highlighted in Zaza et al. (2018), historical analysis of IPM data is necessary to understand good practices, and discover trends and farming behaviors. Therefore, in order to define an efficient tool for the historical analysis of IPM farm data, Zaza et al. (2018) proposed the usage of a dedicated *Data Warehouse* (DW) and the *On-Line Analytical Processing* (OLAP) system, as a decision-making tool of FMISs. In particular, in Zaza et al. (2018), the authors have provided a Business Intelligence tool, called BI4IPM, which could be conveniently used to verify the compliance of the farm operations with the requirements included in the olive crop IPM technical specification provided by the Apulia region. By means of BI4IPM, decision-makers can answer to queries like this: "*What are sustainable practices of each farm of Apulia region during 2017?*".

However, this work suffers of an important limitation: the OLAP model does not take into account climatic data. In this way it is difficult to compare IPM data over different campaigns and farms (which can be geographically distributed over a large area), since time and space can be associated to different climate, and therefore with important impacts on cultures, as above described. For example, it is not possible to answer decisional queries like this: *"What are sustainable practices of each farm of Apulia region during 2017 that had the same climatic conditions?"*

Thus, considering the mandatory and very important relation between climate data on crops and pests, in this paper we extend our previous work (Zaza et al., 2018) integrating climatic data into OLAP model. In particular, we use the Growing Season Index (GSI), a phenology model, to indirectly compare the farms by a climatic point of view. Indeed, phenology is used as an indicator to evaluate the effects of climate change in short and long periods, because it is driven by environmental variables, such as temperature, precipitation and photoperiod (Xu et al., 2014). The GSI is a model developed for the prediction of plant phenology in response to low temperatures, evaporative demands and photoperiod, applicable at global scale (Jolly et al., 2005). In this work, we use a modified version of the GSI proposed by Orlandi et al. (2013). The main difference is that the "original" GSI was developed and tested for the assessment of canopy foliar dynamics on different vegetal species at global scale, while the second version is adapted to the olive trees and tested in several locations in the Mediterranean basin. The integration of the GSI data into the OLAP model of Zaza et al. (2018) has been achieved by a new advanced OLAP model, which has obtained using the adaptation of a new DW design methodology proposed by Sautot et al. (2015). Indeed in our application, the usage of existing DW design methodologies outputs an OLAP model that is not enough expressive to allow decision-makers to query IPM data and correlate them to climatic data.

Moreover, since GSI data is usually visualized using graphical data, in this paper we present a new OLAP client (i.e. the OLAP visualization tool) that is able to represent OLAP data by means of pivot tables for numerical data and also for graphic displays. This new OLAP client allows the simple at glance visual analysis of IPM and GSI data.

We have validated our approach using some real data concerning farms of the Bari department, and open climatic data provided by Apulia Regional Agency for the Prevention and Protection of the Environment (ARPA-Puglia, 2017).

The paper is organized as follow: Section 2 presents a brief background on DW and OLAP approach, IPM context at Apulia Region scale and GSI model; Section 3 is dedicated to the definition of the Apulian case study, exploring the climatic and farming data adopted; Section 4 details the OLAP model for agro-phenological data; Section 5 presents a set of OLAP advanced analysis achieved; Section 6 provides an overview regarding the implementation of the OLAP system; Section 7 explores the related works regarding the phenology models and OLAP for agricultural data; Section 8 concludes the paper and provides the future perspective.

2. Background

In this section, we present main concepts of technologies and methods adopted in this work: Section 2.1 presents Data Warehouse and OLAP systems, Section 2.2 describes an overview of Integrated Pest Management for olive trees, and Section 2.3 presents details of GSI model.

2.1. Data Warehouse and OLAP

Data Warehouse (DW) and OLAP systems are Business Intelligence technologies allowing online analysis of massive volume of multidimensional data. Warehoused data are stored according to the multidimensional model (Gallo et al., 2010; Kimball and Ross, 2013). Data are organized in dimensions and facts. Dimensions are represented by the analysis axes and are organized into hierarchies (for example, cities, departments and regions). Facts are represented by the analysis subjects and are described by numerical attributes called measures (for example, quantity of sold products). Measures are explored with the OLAP operators, which allow for navigating into the DW. Common operators include Slice, which allows selection of a subset of warehoused data, and Drill, which allows for navigating into hierarchies aggregating measure values with SQL aggregation operators (i.e., MIN, MAX, SUM, AVG, etc.). For example, in a retail application, a OLAP query could be "*What is the total quantity of sold products by year, product and region?*".

Usually OLAP models are derived from existing data sources using three different methodologies: data-driven, user-driven and mixed methodologies (Romero and Abelló, 2009). Data-driven methodologies exclusively use the data sources schema to derive OLAP models. User-driven methodologies take into account needs analysis of decision-makers, and mixed methodologies combine data and user driven methodologies. It has been verified in several works that mixed methodologies are the most effective ones to conceive useful OLAP models.

2.2. Integrated pest management

In Italy, the development of the IPM guidelines is assigned to the regions, which take into account their own agro-environmental conditions, the local crops and the related pests. The correct application of the IPM requirements guarantee to farms the improvement of the environmental and economic performances. Moreover, in order to increase the spread of certified integrated production some Italian regions, such as the Apulia Region with the Measure 10.1.1 of the Rural Development Programme (RDP) 2014–2020, have provided an economic incentive for the farmers who decided to adhere to the IPM technical specifications. To get incentives, farmers must meet all the requirements included in the crop specific disciplinary provided by the Region. Furthermore, the compliance between all the farms' operations and the IPM rules is verified by an independent control authority.

The Apulia Region crop's IPM technical specification consist of two parts:

1. *General:* a set of recommendations regarding all the aspects of crop management, such as fertilization restrictions, soil management, and irrigation frequency.
2. *Defense rules:* all the allowed treatments (agronomic and chemical) to protect the crop from the pests listed (Table 1).

The general part of the Apulian technical specifications is made up by all the agronomic requirements that farmers must respect during the whole production cycle, from planting to harvesting. In particular, all

Table 1
Excerpt of the IPM defense rules for the protection of the Olive crop in the Apulia region.

Pest	Intervention criteria	Active substance and auxiliary	Maximum numb. of intervention	Usage limitations and notes
Olive fruit fly (*Bactrocera oleae*)	Intervention threshold:	*Beauveria bassiana*	/	
	• Table olives: presence of the first punctures	Dimethoate	2	
	• Oil olives: 10–15% presence of active infestation (sum of eggs and larvae)	Imidacloprid	1*	* allowed only after flowering
		Fosmet	2	
Olive Moth (*Prays oleae*)	Intervention threshold:	Deltamethrin	1	
	• Table olives: 5–7% presence of active	Fosmet	2	
	infestation (sum of eggs and larvae)	Dimethoate	2*	*allowed only since 8th May 2017 to 4th September 2017
	• Oil olives: 10–15% presence of active infestation (sum of eggs and larvae)	*Bacillus thuringiensis*	/	

the indications provided in the general part of the technical specifications are aimed at improving the phytosanitary condition of the crops, reducing the environmental impact. For example, the fertilization practices must guarantee high quality and quantity of the production, without excess, in order to maintain the soil fertility, protect the environment and prevent pests.

The second part of the IPM technical specification is dedicated to the defense rules (Table 1).

Table 1 shows an excerpt of the IPM defense rules for the Olive crop protection taken from the Apulia region technical specification. In the following, we describe each column:

- *Pest:* This is a list of the pests of the Olive crop. In the example, there are the *Bactrocera oleae* (Rossi) and *Prays oleae* (Bernard), two of the most dangerous pests for olive trees, in the Mediterranean Basin (Hegazi et al., 2007; Tzanakakis, 2006).
- *Intervention criteria:* It represents the threshold before the pest causes economic losses to the cultivated crop. In the example, the value for the table olive cultivars is lower than oil olive cultivars, both for olive fruit fly and olive moth. For the table olives, in the case of *B. oleae*, a single puncture strongly reduces the value of the product, so it is necessary to control the pest immediately, while for the *P. oleae* the damage is less, so a moth presence lower than 5–7%, as a function of different cultivars, is economically acceptable and the control starts when the pest pressure overheads this value. On the other hand, in oil olives the intervention threshold is fixed between 10 and 15%, due to the lower economic damage caused by the two insects.
- *Active substance and auxiliary:* It reports the list of allowed Active Substances (ASs) and auxiliaries that could be used to control the pest. In the example, to control olive moth, phytosanitary products containing *Bacillus thuringiensis*, an entomopathogenic bacteria, Dimethoate, Fosmet or Deltametrhin are authorized.
- *Maximum number of treatments:* It represents the maximum number of treatments admitted for a specific molecule, independently by the controlled pest. For example, products based on *B. thuringiensis* and *Beauveria bassiana* are not limited, because they are natural organisms without environmental impact. In contrast, Dimethoate and Fosmet are limited to two treatments per year, while Imidacloprid and Deltamethrin are restricted to one treatment per year, due to the toxicity of the molecules.
- *Usage limitations and notes:* It is the list of eventual other restriction for specific molecules, such as temporal restrictions. For example, Dimethoate products are allowed to control olive moth only since 8th May 2017 to 4th September 2017, while they are always admitted for treatment against olive fruit fly.

2.3. GSI model

As described by Easterling et al. (2007) higher temperatures and altered patterns of precipitation can probably reduce yields of many crops. For example, Gregory et al. (1999) have reported yields reduction in rice of about 5% per °C rise overhead 32 °C. However, at the same time, the abovementioned environmental factors affect the growth and diffusion of pathogens and pests, such as fungi, insects, bacteria and so on. For example, the presence of the olive fruit fly, one of the most dangerous pest for olive trees in the Mediterranean Basin and the Middle East (Tzanakakis, 2006), is strictly correlated to abiotic environmental conditions (McFadden et al., 1977; Tzanakakis, 2006; Burrack and Zalom, 2008). Absence of olives through the spring and the beginning of summer combined with adverse environmental conditions like high or low temperatures and short photoperiods could cause a facultative reproductive dormancy in *B. oleae* adults (Fletcher et al., 1978; Crovetti et al., 1982). Wang et al. (2009) have highlighted the effects of high temperature on *B. oleae*. In particular, when temperatures rise above 30 °C, the *B. oleae* adults are frenetically active and oviposition is inhibited, while at 35 °C the activity stops (Avidov, 1954). In addition, the larval mortality increases at temperature higher than 30 °C, particularly throughout young larvae (Tsitsipis, 1977), while at 35 °C no flies develop to adult (Genç and Nation, 2008).

In this section, we describe the Growing Season Index for Olive tree we have used in our proposal.

The Growing Season Index (GSI) is a model developed for the prediction of plant phenology in response to low temperatures, evaporative demands and photoperiod, applicable at global scale (Jolly et al., 2005). The GSI can predict not only the beginning and the ending of growing season, but also the canopy status during the year without *a priori* knowledge of the vegetation or climate. In particular, for each variables minimum and maximum threshold limits have been defined, assuming that phenological activity varied linearly from inactive (0) to unconstrained (1).

The low temperatures influence many biochemical processes (Levitt, 1980). For example, water in the xylem of some trees can freeze at temperatures below −2 °C (Zimmermann, 1964). The selected range varied between a lower minimum temperature threshold of −2 °C (*TMMin*) and an upper threshold of 5 °C (*TMMax*), according with Larcher and Bauer (1981). The minimum temperature index (iTMin), has been created as follows:

$$iT_{Min} = \begin{cases} 0 & if\ \ T_{Min} \leqslant T_{MMin} \\ \frac{T_{Min}-T_{MMin}}{T_{MMax}-T_{MMin}} & if\ \ T_{MMax} > T_{Min} > T_{MMin}, \\ 1 & if\ \ T_{Min} \geqslant T_{MMax} \end{cases} \tag{1}$$

where *iTMin* is the daily indicator for minimum temperature bounded between 0 and 1; *TMin* is the observed daily minimum temperature in Celsius degrees; $TM_{Min} = -2$ °C and $TM_{Max} = 5$ °C.

Partial to complete stomatal closure (Mott and Parkhurst, 1991), leaf development rate reduction (Salah and Tardieu, 1996), induction of the shedding of leaves (Childes, 1988), and slowing down or stopping cell division (Granier and Tardieu, 1999) could be caused by water stress. To estimate the water needs, Jolly et al. (2005) selected an index of the evaporative demand, the atmosphere Vapor Pressure Deficit (VPD). According to the literature, they set the VPD_{min} at 900 Pa, because VPD values less than should exert little effect on stomata, while upper threshold VPD_{max} has been fixed at 4100 Pa, because generally greater values of VPD are sufficient to force complete stomatal closure, even when the soils are moist (Osonubi and Davies, 1980; Tenhunen et al., 1982). The VPD index (*iVPD*) was derived as follows:

$$iVPD = \begin{cases} 0 & \text{if } VPD \geqslant VPD_{Max} \\ 1-\frac{VPD - VPD_{Min}}{VPD_{Max} - VPD_{Min}} & \text{if } VPD_{Max} > VPD > VPD_{Min} \\ 1 & \text{if } VPD \leqslant VPD_{Min} \end{cases} \quad (2)$$

where *iVPD* is the daily indicator for VPD limited between 0 and 1; *VPD* is the observed daily VPD in Pascals; $VPD_{Min} = 900$ Pa and $VPD_{Max} = 4100$ Pa.

The photoperiod is a consistent annual climatic indication for plant, because it does not change during the years at a specific location. Many authors have exposed that the photoperiod is strictly connected to both leaf flush and leaf senescence all over the world (Njoku, 1958; Rosenthal and Camm, 1997; White et al., 1997; Häkkinen et al., 1998; Partanen et al., 1998; Borchert and Rivera, 2001). Moreover, a strong interaction between photoperiod and temperature that could limit foliar phenology exist, and temperature changes are ineffective without corresponding photoperiod changes (Partanen et al., 1998). The authors set up the minimum threshold at 10 h, because they assumed that equal or lesser values completely limited canopy development, while the upper limit was fixed at 11 h, assuming that equal or higher values allow canopies to develop unconstrained. The photoperiod index (*iPhoto*) was calculated as follows:

$$iPhoto = \begin{cases} 0 & \text{if } Photo \leqslant Photo_{Min} \\ \frac{Photo - Photo_{Min}}{Photo_{Max} - Photo_{Min}} & \text{if } Photo_{Max} > Photo > Photo_{Min} \\ 1 & \text{if } Photo \geqslant Photo_{Max} \end{cases} \quad (3)$$

where *iPhoto* is the daily photoperiod indicator bounded between 0 and 1; *Photo* is the daily photoperiod in seconds; $Photo_{Min} = 36000$ s (10 h) and $Photo_{Max} = 39600$ s (11 h).

The product of the individual daily indicators for minimum temperature, VPD, and photoperiod forms a single metric, the GSI, which is a daily indicator of the relative constraints to foliar canopy development or maintenance due to climatic limits. GSI is continuous and restricted between 0 (inactive) and 1 (unconstrained). The daily metric (*iGSI*) is calculated as follows:

$$iGSI = iT_{Min} \times iVPD \times iPhoto \quad (4)$$

The daily GSI is then calculated as the 21-day moving average of the daily *iGSI*, to avoid reaction to short-term changes in environmental conditions (Lieberman, 1982).

Orlandi et al. (2013) have presented a new version of the GSI proposed by Jolly et al. (2005). The main difference is that the "original" GSI was developed and tested for the assessment of canopy foliar dynamics on different vegetal species at global scale, while the second version is adapted to the olive trees and tested in several locations in the Mediterranean basin.

The variables taken into account by Orlandi et al. (2013) to adapt the model are explained in the following.

As for the original GSI, iT_{min} is derived by Eq. (1), but T_{MMin} is set up on 0 °C and T_{MMax} is fixed at 7 °C (relation 5).

$$iT_{Min} = \begin{cases} 0 & \text{if } T_{Min} \leqslant T_{MMin} \\ \frac{T_{Min} - T_{MMin}}{T_{MMax} - T_{MMin}} & \text{if } T_{MMax} > T_{Min} > T_{MMin} \\ 1 & \text{if } T_{Min} \geqslant T_{MMax} \end{cases} \quad (5)$$

where *iTMin* is the daily indicator for minimum temperature bounded between 0 and 1; *TMin* is the observed daily minimum temperature in Celsius degrees; $TM_{Min} = 0$ °C and $TM_{Max} = 7$ °C.

Moreover, *iPhoto* is calculated with the same relation (3) and thresholds proposed by Jolly et al. (2005).

In this version of GSI, the water requirements are estimated with the potential evapotranspiration (ETP) instead of VPD. Therefore, the ***iETP***, that substitutes ***iVPD***, derived by the following relation:

$$iETP = \begin{cases} 0 & \text{if } ETP \geqslant ETP_{Max} \\ 1-\frac{ETP - ETP_{Min}}{ETP_{Max} - ETP_{Min}} & \text{if } ETP_{Max} > ETP > ETP_{Min} \\ 1 & \text{if } ETP \leqslant ETP_{Min} \end{cases} \quad (6)$$

where ***iETP*** is the daily index for ETP bounded between 0 and 1; ***ETP*** is the daily evapotranspiration value (mm d^{-1}) estimated with the Priestley–Taylor method (Priestley and Taylor, 1972); $ETP_{Min} = 2$ mm d^{-1} and $ETP_{Max} = 5$ mm d^{-1}.

The daily solar radiation is also included in the Mediterranean GSI. The daily index ***iRad*** is obtained with the following relation:

$$iRad = \begin{cases} 0 & \text{if } Rad \leqslant Rad_{Min} \\ \frac{Rad - Rad_{Min}}{Rad_{Max} - Rad_{Min}} & \text{if } Rad_{Max} > Rad > Rad_{Min}, \\ 1 & \text{if } Rad \geqslant Rad_{Max} \end{cases} \quad (7)$$

where ***iRad*** is the daily index for the solar radiation included between 0 and 1; ***Rad*** is the daily solar radiation value (MJ m^{-2} d^{-1}) estimated with the Campbell and Donatelli model (Donatelli and Campbell, 1998); $Rad_{Min} = 5$ MJ m^{-2} d^{-1} and $Rad_{Max} = 12$ MJ m^{-2} d^{-1}.

Finally, the daily metric ***iGSI*** is calculated with the product of the new variables:

$$iGSI = iT_{Min} \times iPhoto \times iETP \times iRad \quad (8)$$

Again, the daily GSI is calculated as the 21-day running average of the daily *iGSI*.

An example of the adapted GSI for the Apulian city of Foggia taken by Orlandi et al. (2013) is shown in Fig. 1.

As shown in Fig. 1, the GSI obtained for the city of Foggia presents two active period, the first during the spring, from February to June, the second during the autumn, from September to November and highlighted by the two pikes of phenological activity. During the summer, from June to September, the ETP conditions inactivate the phenological activity, as underlined in Fig. 1 by the red area.

3. Case study

In this section we introduce the data we have used to validate our proposal. In particular, Section 3.1 presents climatic data, and Section 3.2 describes farming data.

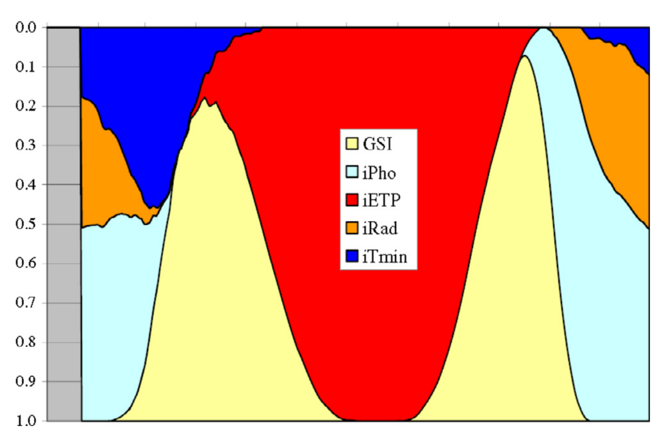


Fig. 1. Daily index for Foggia, as minimum temperature (iTmin; blue), solar radiation (iRad; orange), photoperiod (iPho; light blue), evapotranspiration (iETP; red), and growing season (GSI; yellow). The seasonal limits on GSI of each variable are shown. The GSI is represented as the complement to 1. From Orlandi et al. (2013). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Excerpt of ARPA-Puglia climatic data on Bari.

Date	Time	TMIN (°C)	TMAX (°C)	Rain (mm)
03/01/2017	00:00:00	7.0	7.3	0.0
03/01/2017	00:30:00	6.7	7.1	0.0
03/01/2017	01:00:00	7.0	7.6	0.0
03/01/2017	01:30:00	7.0	7.6	0.0
03/01/2017	02:00:00	6.6	7.2	0.0
03/01/2017	02:30:00	6.2	6.7	0.0
03/01/2017	03:00:00	6.3	7.0	0.0
03/01/2017	03:30:00	6.7	7.7	0.0
03/01/2017	04:00:00	7.7	9.1	0.0
03/01/2017	04:30:00	9.1	9.5	0.0
03/01/2017	05:00:00	9.5	10.0	0.0
03/01/2017	05:30:00	9.7	10.0	0.0
03/01/2017	06:00:00	9.1	9.8	0.0
03/01/2017	06:30:00	8.7	9.2	0.0
03/01/2017	07:00:00	8.6	8.9	0.0
03/01/2017	07:30:00	8.7	9.2	0.0
03/01/2017	08:00:00	9.0	9.2	0.0
03/01/2017	08:30:00	9.0	9.5	0.0
03/01/2017	09:00:00	9.2	9.3	0.0
03/01/2017	09:30:00	9.2	9.7	0.0
03/01/2017	10:00:00	9.6	10.7	0.0
03/01/2017	10:30:00	10.8	11.3	0.0
03/01/2017	11:00:00	11.2	11.6	0.0
03/01/2017	11:30:00	11.5	13.1	0.0
03/01/2017	12:00:00	13.0	13.7	0.0
03/01/2017	12:30:00	13.5	13.9	0.0
03/01/2017	13:00:00	13.5	13.9	0.0
03/01/2017	13:30:00	13.4	13.9	0.0
03/01/2017	14:00:00	13.1	13.5	0.0
03/01/2017	14:30:00	13.0	13.4	0.0
03/01/2017	15:00:00	11.9	13.0	0.0
03/01/2017	15:30:00	11.2	12.0	0.0
03/01/2017	16:00:00	10.2	11.2	0.2
03/01/2017	16:30:00	9.6	10.3	0.2
03/01/2017	17:00:00	9.5	9.9	0.0
03/01/2017	17:30:00	9.4	9.5	0.0
03/01/2017	18:00:00	9.4	9.6	0.0
03/01/2017	18:30:00	9.5	9.7	0.0
03/01/2017	19:00:00	9.5	9.8	0.0
03/01/2017	19:30:00	9.5	9.8	0.0
03/01/2017	20:00:00	9.5	9.6	0.0
03/01/2017	20:30:00	9.6	9.8	0.0
03/01/2017	21:00:00	9.4	9.8	0.0
03/01/2017	21:30:00	9.0	9.5	0.0
03/01/2017	22:00:00	8.8	9.1	0.0
03/01/2017	22:30:00	8.7	8.8	0.0
03/01/2017	23:00:00	8.5	8.7	0.0
03/01/2017	23:30:00	7.9	8.4	0.0

3.1. Climatic data

The climatic data are extracted from the ARPA-Puglia network (ARPA-Puglia, 2017). The raw data taken from ARPA-Puglia are measures recorded every 30 min for Minimum and Maximum Temperature (°C) and the amount of rain in mm (Table 2).

An excerpt of the dataset taken from ARPA-Puglia, related to the 3rd January 2017 is presented in Table 2.

3.2. Farming data

The phytosanitary data used in this work are obtained from four real Apulian farms located in the Bari department and renamed for privacy reason. The data extracted from the farms 1 and 3 are related to the 2016–2017 campaign, while for the farm 4 the data are related to the campaign 2017–2018. Finally, the data taken from the farm 2 are related to the two analyzed campaigns, 2016–2017 and 2017–2018. An example of phytosanitary data taken from the farm 2 and related to the campaign 2017–2018 is shown in Table 3.

Table 3
Excerpt of IPM data.

Date	Plot	A. S.	Pest
15/03/2017	P1_2	Dodine	*S. oleagina*
28/04/2017	P1_2	Trifloxystrobin -Tebuconazolo	*C. gloesporioides*
13/06/2017	P1_2	Dimethoate	*P. oleae*
12/09/2017	P1_2	Copper	*S. oleagina*

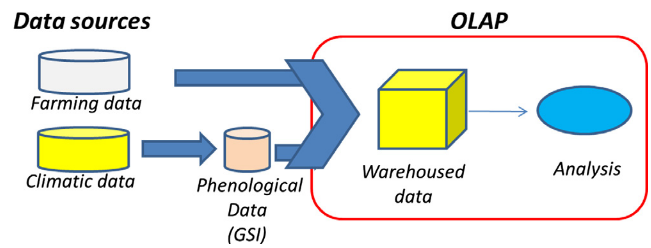


Fig. 2. Generic framework for the design of OLAP model integrating climatic and farming data.

4. OLAP model for agro-phenological data

As discussed in Section 2.1, mixed DW design methodologies take into input data sources and user needs and they output an OLAP model, and are the most used for real-life OLAP applications.

Then, in this section, we present two OLAP models for the integration of GSI and IPM data.

The generic framework used to achieve this integration is described in Fig. 2. The main idea is to use data sources containing farming and climatic data. Then, climatic data are used to obtain phenological data using the GSI model. Finally, the OLAP model is derived from farming and phonological data. The warehoused data is then explored using an OLAP client.

Following this approach, the Section 4.2 is dedicated to the OLAP model. Therefore, in the Subsection 4.2.1 we present a naïve approach based on existing DW design methodologies, and our new model that allows an effective agro-climatic multidimensional analysis presented in Subsection 4.2.2. In Section 4.1 we describe how we have obtained the GSI data.

4.1. Apulian GSI distributions

To calculate the GSI related to the city of Bari, for each day, the lowest (T_{min}) and the highest (T_{max}) values of temperature are extracted, while for the amount of rain all the single measurements are summed. Therefore, in the provided example (Table 2), the data extracted are T_{min} = 6.2 °C, T_{max} = 13.9 °C and *Rain* = 0.4 mm. The mean imputation method (Acuna and Rodriguez, 2004), at month scale, is used for the treatment of the missing value. The obtained daily T_{min}, T_{max} and *Rain* are used to feed the software RadEst version 3.00 (Donatelli et al., 2003), in order to estimate the daily solar radiation (*Rad*) and potential evapotranspiration (*ETP*), according to Orlandi et al. (2013). In particular, the solar radiation data are estimated by the Campbell-Donatelli model (Donatelli and Campbell, 1998), while the Priestley–Taylor method (Priestley and Taylor, 1972) is used to estimate the ETP values. Finally, the daily photoperiod (*Photo*) for the city of Bari is estimated by the latitude and yearday (Monteith and Unsworth, 1990).

The obtained daily values *Photo*, T_{min}, *ETP* and *Rad* are used to calculate *iPhoto* (3), iT_{min} (5), *iETP* (6) and *iRad* (7), respectively. Then, the daily metric *iGSI* is calculated with the Eq. (8) of Section 2.3. Finally, the daily GSI is the result of the 21-day running average of the daily *iGSI*. In Fig. 3 the obtained GSI daily distribution for the 2016 and the 2017 are presented.

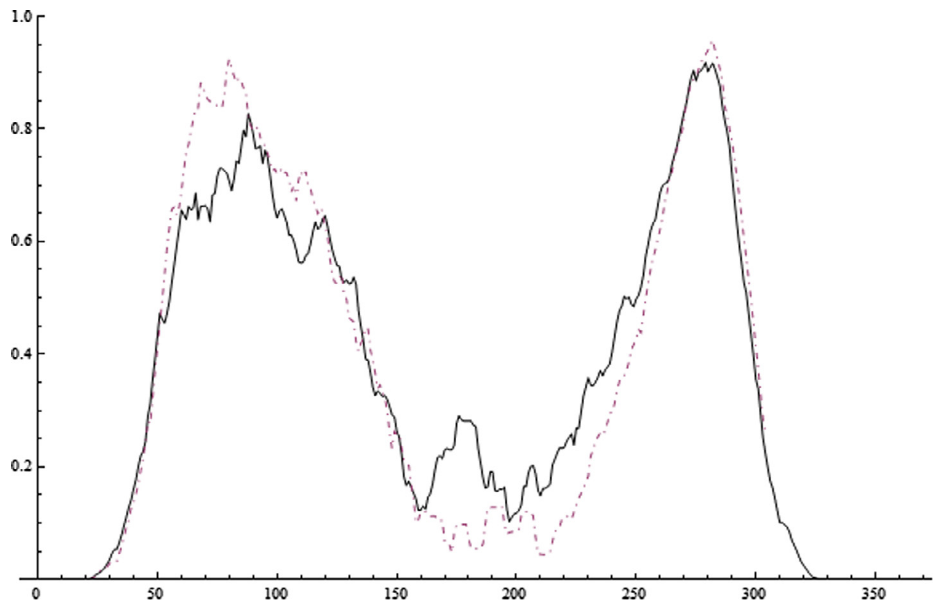


Fig. 3. GSI daily distribution for Bari. The continue line represents the GSI for the year 2016, while the dot-dashed line represents the 2017 GSI distribution.

The obtained GSI distributions are tested by the Cramér-von Mises normality test. In our case the hypothesis that the GSI distributions for Bari 2016 and Bari 2017 (Fig. 3) are distributed according to the Normal Distribution is rejected at 5% level (p-value = 0).

The phenological activity obtained for the city of Bari starts from February to November, with pikes in March and October (Fig. 3). Moreover, in 2016 a third pike occurred during June that is not present in 2017. Comparing the obtained GSI distributions of Bari with the reported GSI distribution of Foggia provided by Orlandi et al. (2013), a difference is detected in the summer period. In fact, the phenological activity from June to September in Foggia is arrested, while in Bari the activity decreases, but it is never stopped, for both 2016 and 2017.

4.2. OLAP model

4.2.1. Naive OLAP model

Based on the existing mixed driven methodology (Romero and Abelló, 2009), using GSI and farming data described in Section 3, we obtained the OLAP model of Fig. 2.

This multidimensional model (Fig. 4) presents two related facts (i.e. constellation model).

The fact IPM allows the analysis of the IPM defense rules, and it is described by five dimensions:

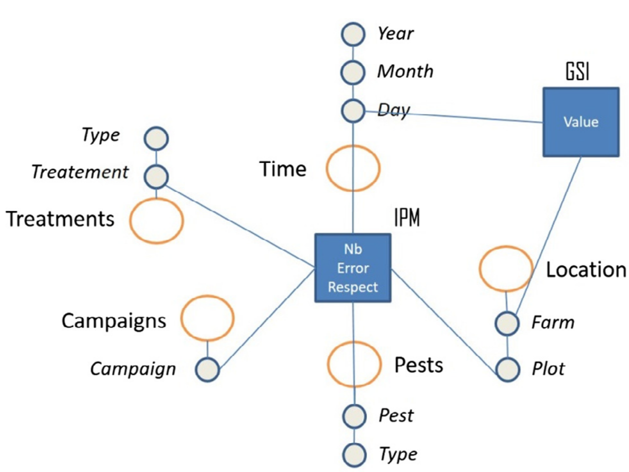


Fig. 4. Olive crop IPM conventional multidimensional model.

- *Time:* The temporal dimension is composed by the hierarchy day, month, and year.
- *Campaign:* It represents the crop campaign (for example 2016–2017).
- *Location:* It describes the organization of the farms, grouping the plots in farm, in order to allow analysis of the performed treatments at different spatial scales. For example, the farmer could treat two plots in different ways during a single campaign in order to evaluate the best management strategy to apply for the next crop cycle.
- *Pests:* This dimension represents the pests listed in the IPM crop technical specification (first column, Table 3). Pests are structured by type, such as weed, fungi, insects, and bacteria. Therefore, each pest is classified by the scientific name, e.g., *Bactrocera oleae*, *Prays oleae*, etc. (Table 3).
- *Treatments:* This dimension includes the types of interventions allowed in the IPM technical specification, organized by type, Agronomic (e.g., Pruning) and Chemical (e.g., Dimethoate, Imidacloprid, etc.) (third column, Table 3).

The measures are as follows:

- *Nb:* This represents the number of performed treatments (for example, the application of Copper during the 2016–2017 campaign). The sum function is used for the aggregation of this measure by all dimensions.
- *Respect:* This Boolean measure defines if the executed intervention adheres to the IPM defense rules. For example, the usage of Deltamethrin to control *B. olea* is not allowed. Therefore, the measure value for this treatment is "NO". The AND operator is used for the aggregation of *Respect* measure.
- *Error Type:* This measure defines the treatment based on the IPM rules. It is an enumeration value described as follows:
  - o *Allowed:* The operation is admitted (for example, to control the fungi *Colletotrichum gloesporioides* with products based on Tebuconazole-Trifloxystrobin).
  - o *A.S. not Allowed:* The treatment is not admitted (for example, the application of products based on Glyphosate 30.8%).
  - o *Exceeded nb inter:* The sum of treatments performed in a year overheads the maximum number of treatments admitted (fourth column, Table 3); for example, a third intervention based on Fosmet product to control *B. oleae* or *P. oleae* is not allowed.
  - o *A.S. not allowed in time:* The treatment executed is not allowed in

Farm	Campaign	Pests	Treatments	Number of Treatments
+Farm2	2016-2017	Spilocea oleagina	Copper	2
		Bactrocera olea	Deltamethrin	1
			Dimethoate	1
			Imidacloprid	1
	2017-2018	Spilocea oleagina	Copper	1
		Prays oleae	Dimethoate	1
−Farm4	2017-2018	Spilocea oleagina	Copper	4
		Prays oleae	Bacillus thuringiensis	2
			Deltamethrin	1
			Fosmet	2
P1_4	2017-2018	Spilocea oleagina	Copper	2
		Prays oleae	Bacillus thuringiensis	1
			Fosmet	1
P2_4	2017-2018	Spilocea oleagina	Copper	2
		Prays oleae	Bacillus thuringiensis	1
			Deltamethrin	1
			Fosmet	1

a)

Time	Farm	GSI
+08/16	Farm1	0,319
−09/16	Farm1	0,665
01/09/2016	Farm1	0,503
02/09/2016	Farm1	0,498
03/09/2016	Farm1	0,501
04/09/2016	Farm1	0,488
05/09/2016	Farm1	0,484
06/09/2016	Farm1	0,498
07/09/2016	Farm1	0,504
08/09/2016	Farm1	0,519
09/09/2016	Farm1	0,54
10/09/2016	Farm1	0,574
11/09/2016	Farm1	0,595
12/09/2016	Farm1	0,619
13/09/2016	Farm1	0,632
14/09/2016	Farm1	0,639
15/09/2016	Farm1	0,662
16/09/2016	Farm1	0,686
17/09/2016	Farm1	0,699
18/09/2016	Farm1	0,703
19/09/2016	Farm1	0,705
20/09/2016	Farm1	0,715
21/09/2016	Farm1	0,737

b)

Time	GSI	Number of Treatments
15/09/2016	0,662	5
16/09/2016	0,686	2
28/09/2016	0,871	1
30/09/2016	0,904	1
11/10/2016	0,876	4
14/10/2016	0,796	2
15/10/2016	0,769	2
25/10/2016	0,397	1
15/03/2017	0,842	1
11/04/2017	0,721	2
28/04/2017	0,66	1
05/05/2017	0,526	2

c)

Fig. 5. (a) Example of query performed on the IPM fact; (b) example of query performed on the GSI fact; (c) example of query performed thanks to the Drill-Across operator.

that period (fifth column, Table 2). For example, interventions based on Dimethoate products to control *P. oleae* are forbidden before 8th May 2017 and after 4th September 2017.

- o *Errors.* This appears when there are more errors. A drill-down operation is required to recover the detailed error type occurred for a specific operation.

A user-defined aggregation is executed for *Error type*, displaying *Errors* when more than one of *A.S. not Allowed* OR *Exceeded nb inter* OR *A.S. not allowed in time* are aggregated; otherwise, the model output is *Allowed.*

The other fact GSI represents the daily value of the GSI (*value* measure), and it is described by the two dimensions *Time* and *Location*. The only difference with the IPM fact is that the Location dimension is associated at the Farm level.

Using this constellation model is possible to visually querying each fact for independently exploring IPM and GSI data. An example of these two kinds of queries is shown in Fig. 5. Fig. 5a shows the OLAP query "*Show the number of performed treatments for Farm2 and Farm4 organized by campaign, pests and treatments*" on the IPM fact.

Fig. 5b shows the OLAP query "*What is the September daily GSI value calculated for the Farm1?*" on the GSI fact.

The comparison of the two data sets must be hand done by the decision-makers at glance, which is difficult and sometime impossible (when data is huge).

To solve this problem, OLAP systems provide the Drill-Across operator. This operator allows to analyze in the same OLAP client facts uniquely using common dimensions, and grouping to "All" the other dimensions. In our case study, this means that we can exclusively analyze the IPM measures with the GSI value according to the *Time* dimension (*Location* is not common since they have a different detailed level). Other dimensions cannot be used for the OLAP analysis. Fig. 5c shows the results of the OLAP query "*Show the days where one or more treatments occurred and the related daily GSI value*".

This query is not useful for an effective analysis, since pests, treatments and farms cannot be visualized in the same pivot table.

This problem is issued from the used DW design methodology, which does not allow using factual data to analyze another fact in a constellation model.

Therefore, to solve this problem we have adapted the new DW design methodology proposed by Sautot et al. (2015) as described in the next subsection.

4.2.2. Advanced OLAP model

Sautot et al. (2015) present a mixed multidimensional refinement methodology that transforms a constellation schema by defining hierarchy levels using a hierarchical clustering algorithm. This refinement methodology enriches a dimension with factual data. In particular, the methodology notes "source fact" the fact that is removed and used to enrich the "target dimension" with new hierarchies created using a data mining algorithm.

The proposed model is shown in Fig. 6.

In our work, we have adopted this methodology using the GSI fact as source fact. The target dimension is the *Location* dimension. But, to meet our decision-maker analysis needs we provide three modifications to Sautot et al. (2015):

1. The measure of the "source fact" (i.e. *value of GSI*) are added to the other fact (i.e. *IPM*)
2. A new dimension is added with the new hierarchies (i.e. *Similarity*)
3. The data mining algorithm is replaced by a statistical method, the *U* test (Mann and Whitney, 1947), to compare the GSI distributions. This test allows evaluating if two not normal distributions are significantly different. In our example, the test returns that the two distributions are significantly different at 5% level (p-value = 0.019). According to the *U* test result, the GSI distributions, expression of the climatic factors, are significantly different for the farm in the Bari department for the year 2016 and 2017.

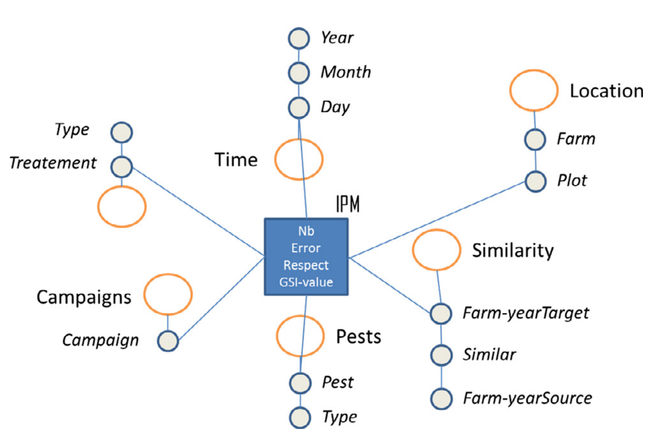


Fig. 6. Olive crop IPM advanced multidimensional model.

The resulting OLAP model is shown in Fig. 6.

Let us describe the new *Similarity* dimension, which is obtained to our statistical method previously described.

The *Farm-yearSource* level represents agro-climatic conditions of farm by year, for example *Farm2-2017* represents the GSI condition of the Farm2 during 2017, *Farm1-2016* represents the GSI condition of the Farm1 during 2016, etc.

The *Similar level* groups all other farms-year in *similar* and *not similar*, based on the result of the Mann-Whitney test performed on the GSI distribution assigned to the selected farm and the others loaded in the system.

Finally, *Farm-yearTarget* represents the farms similar or not, respect to the chosen farm (*Farm-yearSource*). For example the farms that are not similar to *Farm1-2016* are: *Farm1-2017* and *Farm4-2017*.

An example of this hierarchy is shown in Fig. 7.

Using our multidimensional advanced model it is possible to answer to OLAP queries using at the same time GSI and IPM measures, and exploring farms that are similar according to the GSI value. An example is shown in Fig. 8. It shows the average of treatments at plot level (*AVG_Treatments*) over 2017 and 2016 campaigns and the GSI values, for the farms not similar and similar to the Farm1 in the campaign 2016. In particular, as shown in Fig. 8, the advanced multidimensional model allows performing a set of queries over different farms, based on the similarity obtained by the comparison of the GSI distributions. In the provided example (Fig. 8), the user farm, *Farm2_2016*, is compared with similar, *Farm1_2016* and itself, and not similar farms, *Farm2_2017* and *Farm4_2017*, according to *Similar* dimension (Fig. 7), previously described. The similarity or not, is underlined in the system by the graphic representation of the two GSI distributions investigated. In particular, for the similar farms is shown the same GSI annual distribution, while for the not similar farms, the system returns the comparison between the two distributions, highlighting the differences occurred.

To conclude, our new model solves the issues of the previous described conventional OLAP model.

All other analysis possibilities provided by our model are described in the next section.

5. OLAP advanced analyses: results

In this section we present a set of examples that are representative of the analysis possibilities offered by our new OLAP model.

Among all analysis capabilities offered by the model (c.f. (Zaza et al., 2018) for more details), in this work, in particular, we show how it is possible to analyze IPM data of different farms having the same phenological conditions over a year to understand some best practices (Section 5.1).

Moreover, we present some examples of the analysis of the practices, and therefore of the causes, of farms having different climatic conditions during a year (Section 5.2).

5.1. Best practices example

Let us consider the query OLAP shown in Fig. 9. This picture shows that the similar farms *Farm1_2016* and *Farm2_2016* have not respected the IPM defense rules. The proposed system allows focusing on these two similar farms in order to analyze through the measures *Error type* and *Number of treatments*, each performed treatments, taking into account the target pests (Fig. 10). The number of performed treatments is the same, five per farm, but the *Farm1* presents one wrong treatment, while the *Farm2* has two errors. Considering that the two farms are

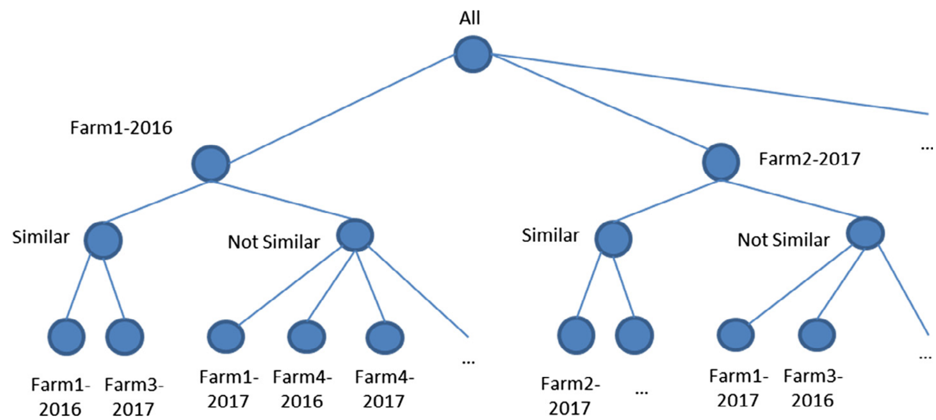


Fig. 7. Example of the similar dimension's hierarchy.

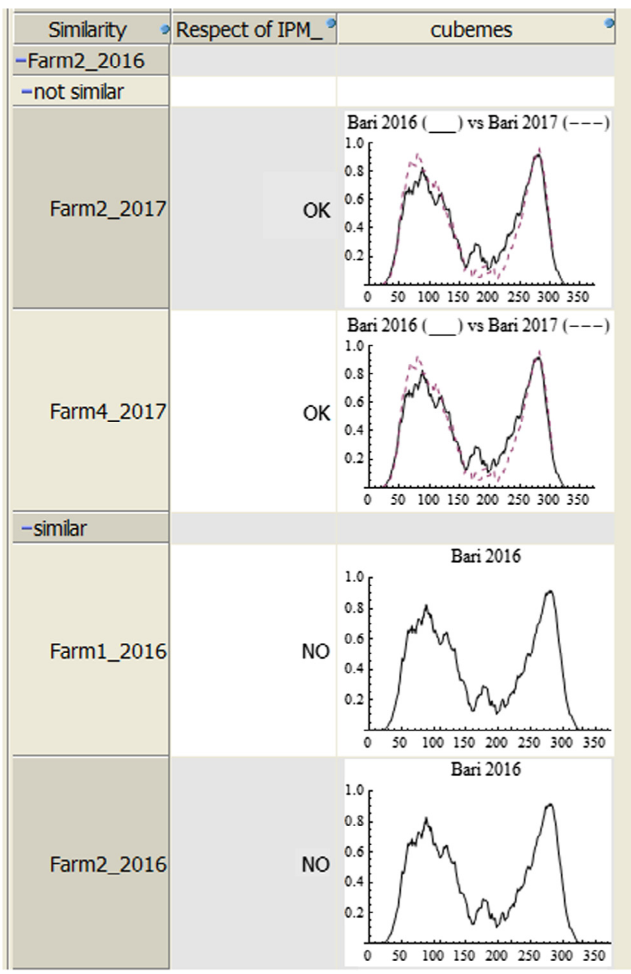


Fig. 8. Example query over IPM and GSI data.

similar by the phenological point of view, the crop protection strategy adopted by *Farm1* could be identified as the more sustainable. Therefore, this type of query could be used to investigate and suggest best practices, taking into account the farms grouped per similar phenological conditions, allowing to improve the economic and environmental performances of the farms recorded in the system.

5.2. Different practices example

As shown in Fig. 9, the *Farm2*, is considered similar for the year 2016 and not similar for the year 2017, respect to *Farm1*. So, in Fig. 11 is presented a query concerning one farm, *Farm2*, aimed at highlighting the differences occurred during the two analyzed years, taking into account the protection versus the insects. It is interesting to point that the phytosanitary treatments to control the insects have changed between 2016, focused on *B. olea*, and 2017, concentrated on *P. olea*.

The different phytosanitary strategies shown in Fig. 11 at level of single farm could be associated to the different climate conditions occurred during the two analyzed campaigns, which have determined the two different GSI distributions, previously presented (Fig. 3). In order to evaluate the hypothesis that the different climate conditions have determined differences in the phytosanitary strategy to control insects, an inter-farm query is performed involving all the farms recorded in the system, taking into account the chemical treatments used to control the *B. olea* and *P. olea*, according with the measures *AVG_Treatments* and *Number of treatments*. As shown in Fig. 12, the number of treatments performed to control the *B. olea* and *P. olea* is completely different between the two years. Indeed, for the 2016 the treatments executed for the *B. olea* is equal to 10 in total (8 considering the plot average), while in 2017 no treatment is performed to control the olive fruit fly. On the contrary, for the 2016 the number of treatments executed for the *P. olea* is equal to 2 in total (1 considering the plot average), while in 2017 to control the olive moth 5 treatments were performed in total (3.5 for the plot average). The showed results seem to confirm the presented hypothesis.

6. Implementation

The OLAP system, based on the advanced model described in Fig. 6, is a three-tier relational OLAP system structured as follows (Fig. 13):

- The Data Warehouse tier. This level is in charge for data storage and it is carried out using the Postgres DBMS. Data are modeled using the star schema logical model (Kimball and Ross, 2013). Star schema denormalizes dimension tables to avoid expensive join operations.

The star schema model implemented in our case study is shown in Fig. 14.

- The OLAP Server tier. This implements the OLAP operators (Drill-Down, Roll-Up, etc.). It has been implemented using Mondrian.
- The OLAP Client tier. It is in charge of the visualization and exploration of warehoused data and is implemented using the OLAP client JRubik. Moreover, the OLAP client has been modified in order to allow the visualization of the graphics representation of the GSI distribution, as shown in the previous section.

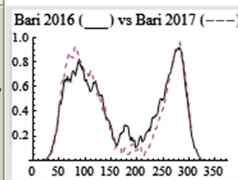
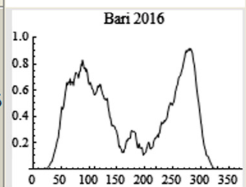
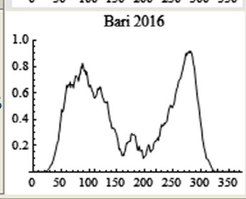
Similarity	cubemes	Error type_	Respect of IPM_	Number of Treatments_
-Farm1_2016				
-not similar				
Farm2_2017		Allowed	OK	4
-similar				
Farm2_2016		Errors	NO	5
Farm1_2016		A. S. not allowed in time	NO	5

Fig. 9. OLAP query exploring similar and not similar farms.

7. Related work

The vegetation phenology studies the recurrence of vegetative cycles in function of climate conditions. Robust phenology models have been used to monitor and predict the vegetation response to climatic data (White et al., 1997). In literature, many works are focused on phenology as an indicator to evaluate the effects of climate change (Kramer, 1996; Lechowicz and Koike, 1995; Schwartz, 1999), because it is driven by environmental variables, such as temperature, precipitation and photoperiod (Xu et al., 2014).

Phenology models have been developed for many crops.

Chunie (2000) developed a phenological model for the budburst of trees, called Unified model. This model was validated and tested for several vegetal species, such as *Aesculus hippocastanum* L., *Buxus sempervirens* L., *O. europaea* L.

García de Cortázar-Atauri et al. (2005) presented a model to calculate the budbreak dormancy in grapevine (*Vitis vinifera*), called BRIN model, tested in different locations in France and for several varieties.

Caffarra and Eccel (2010) developed a phenological model adapted for the grapevine, especially the cultivar Chardonnay. In particular, the authors calibrated and validated several models for the budbreak, flowering and veraison, obtaining a robust phenological model for the cv Chardonnay. The model was validated by environmental data taken from several locations in Italy.

Parker et al. (2011) described a general phenological model to predict the flowering and veraison of grapevine. In particular, the authors tested three different models. Through the optimization and validation of the most accurate by a phenological dataset, the authors have obtained a new general model, tested on 11 cultivars.

Miranda et al. (2013) evaluated several phenological models for the prediction of bud development in peach (*Prunus persica* L. Batsch). For this work the authors have used phenological data acquired for 6 years in different locations in Spain for 3 cultivars.

Rea and Eccel (2006) presented a model for the prediction of the flowering date in apple (*Malus domestica* Borkh.). The model was optimized taking into account 6 phenological series for the cv Golden Delicious at 6 locations in Italian fruit-growing region Trentino.

Recently, the OLAP approach have been used to analyze agricultural data.

Chaudhary et al. (2004) presented a system to analyze the economic factors affecting the yield of cereal crops. The dimensions included in the system are the spatial dimension, the temporal dimension, the fertilizers and the cereals.

Chaturvedi et al. (2008) investigated factors affecting cotton production, defining 13 data cubes related to the analysis of crops taking into account several aspects such as soil and socioeconomic resources using a hybrid design methodology. In this work, the authors have faced the problem related to the multi-granular data, which consists of data sources that are not all available at the same dimension levels, disaggregating data, when possible, using complex interpolation

Similarity	Treatments	Pests	Error type_	Number of Treatments_
Farm1_2016	Copper	Spilocea oleagina	A. S. not allowed in time	2
	Dimethoate	Bactrocera olea	Allowed	2
	Imidacloprid	Bactrocera olea	Allowed	1
Farm2_2016	Copper	Spilocea oleagina	A. S. not allowed in time	2
	Deltamethrin	Bactrocera olea	A. S. not allowed	1
	Dimethoate	Bactrocera olea	Allowed	1
	Imidacloprid	Bactrocera olea	Allowed	1

Fig. 10. OLAP query focus on similar farms.

Similarity	Treatments	Pests	Time	Number of Treatments
Farm2_2017	Dimethoate	Prays oleae	+06/2017	1
Farm2_2016	Deltamethrin	Bactrocera olea	+10/2016	1
	Dimethoate	Bactrocera olea	+09/2016	1
	Imidacloprid	Bactrocera olea	+10/2016	1

Fig. 11. Example of intra-farm query across two years focused on insects phytosanitary treatments.

Pests	Similarity	AVG Treatments_	Number of Treatments_
Bactrocera olea	-not similar		
	Farm2_2017		
	Farm4_2017		
	-similar		
	Farm1_2016	3	3
	Farm2_2016	3	3
	Farm3_2016	2	4
Prays oleae	-not similar		
	Farm2_2017	1	1
	Farm4_2017	2,5	5
	-similar		
	Farm1_2016		
	Farm2_2016		
	Farm3_2016	1	2

Fig. 12. Example of inter-farm query across two years focused on phytosanitary treatments to control insects.

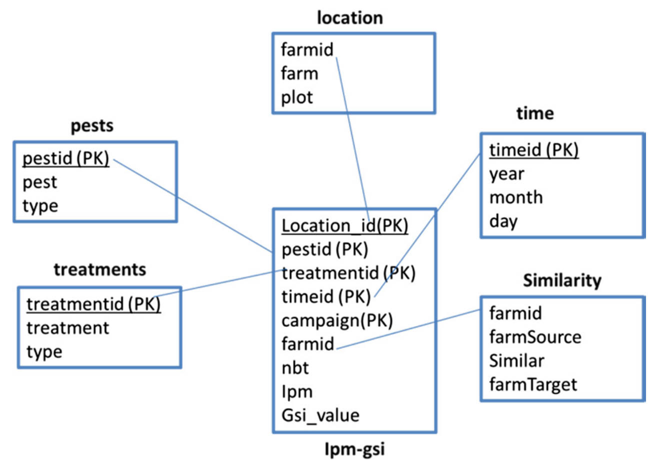


Fig. 14. Implementation of the model of Fig. 6.

functions. The provided data cubes could be explored using not only the classical SQL, but also with some advanced aggregations, such as a weighted average for the market price.

A model to estimate the crop performance for a specific area, weather and environmental domain was developed by Gupta and Mazumdar (2013). The authors defined a data cube able to investigate the success of a specific crop in function of irrigation, weather, seeding and the soil type. Moreover, the schema of the data warehouse is detailed, providing some SQL example decisional queries.

In the work of Pestana et al. (2005) a data model to face the problem of integrating spatial and temporal data providing information is proposed to simplify semantic interoperability and data analysis in a Spatial Data Warehouse (SDW) that uniformly manages all data types to study the vineyard production. The model provides a temporal dimension, a spatial dimension, which represents the plots, and others thematic dimensions. The number of grafts and the number of changes of the geometric shape of the spatial phenomenon are the measures included in the model. Finally, the presented model allows investigating the evolution of the production and the parcel.

Deggau et al. (2010) also focused on farm activities and agricultural production. The proposed system enriches SDWs with ontologies enabling to find the more appropriate SDW by a keyword search.

The support of macro-level planning activities is the aim of the OLAP system developed by Nilakanta et al. (2008), which could be used for the analysis of animal and crop resources. For the design of the multidimensional model, the authors took into account the available data and the decision-maker needs, validating it through a usability test. Therefore, applying a hybrid methodology and considering the data sources and the feedback obtained with the usability test, the authors provided several cubes organized in a constellation schema. The model dimensions are classical, i.e. the product dimension, and complex, such as temporal and spatial dimensions, facing the multi-granular problems as previously described.

The integration of forestry data in a spatiotemporal DW has treated by Miquel et al. (2002), providing a complex multidimensional model, composed by the temporal and the spatial dimensions to measure the forest surface. In this work, the authors reported that the main problem concerns the spatial hierarchy, which is dependant over time.

Pinet and Schneider (2010) adopted the OLAP approach to analyze the usage of agricultural fertilizers. The presented spatio-multidimensional model, based on a UML profile, allows investigating the spread amount per commune and fertilizer. To manage the logical coherence issues due to the heterogeneous data sources, the authors adopted the OCL constraints.

Bimonte et al. (2013a) during the DispEAU project, which aims at monitoring meteorological data collected from French vineyards, evaluated the effect of humidity and temperatures on vineyard production. To enrich the objective of the project was designed a Spatial-OLAP (SOLAP) model presenting a spatial dimension, representing plots and farms, and a temporal dimension. The measures, temperature and

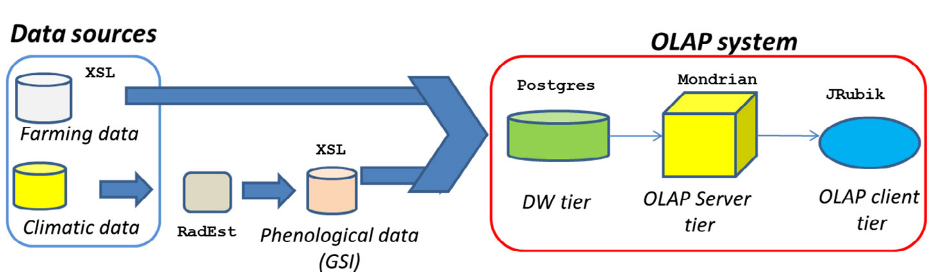


Fig. 13. Architecture of the system.

moisture, are recorded through a wireless sensors network. The model is implemented in an OLAP-GIS architecture using Map4Decision as Web SOLAP client. Adopting a similar SOLAP architecture, the energy consumption of farms was analyzed for the wheat crop (Bimonte et al., 2013b) and the milk production (Bimonte et al., 2014).

8. Conclusion and future work

Olive tree is one of the most important crop at global scale. Apulia is the first region for olive crop in Italy, producing 0.99 Mt, equal to 35% of the national amount (2.81 Mt) (ISTAT, 2016). Therefore, in the Apulia region, a large number (227245) of farms work on the Olive tree (ISTAT, 2010), generating a huge amount of IPM data.

In this paper we present an extended version of the OLAP model proposed in our previous work (Zaza et al., 2018), integrating the olive GSI, a phenology model proposed by Orlandi et al. (2013), in order to indirectly compare the farms by a climatic point of view. Therefore, the GSI for the city of Bari for the years 2016 and 2017 is calculated using the open access climatic data provided by Apulia Regional Agency for the Prevention and Protection of the Environment (ARPA-Puglia). In addition, we charge in the system the phytosanitary data derived from four real Apulian farms located in the Bari department and renamed for privacy reason, related to campaigns 2016–2017 and 2017–2018.

To integrate the GSI distributions and the farming data into the system, we achieved a new advanced OLAP model, adapting a new DW design methodology proposed by Sautot et al. (2015). Moreover, we have introduced a new OLAP client able to represent OLAP data by means of pivot tables for numerical data and for graphic displays, in order to visualize the graphical representation of the GSI distribution. The proposed system allows analysing IPM data of different farms having the same phenological conditions over a year to understand some best practices. The characterization and the spread of the best practices is fundamental to enhance the economic and environmental performances of the farms recorded in the system. At the same time, the system permits to highlight and explain the different practices adopted by farms working in different phenological, and therefore climatic, conditions.

As current work we plan to increase the farming and climatic data, taking into account more farms distributed on different areas, in order to validate the system on a wider territory with bigger datasets. We will also study other crops such as wheat, grapevine, etc. As future work, we will study the usage of sensors to collect climatic data to support precision agro-climatic analysis. To achieve this goal, we plan to deploy our system under big data platforms, such as Hadoop.

References

Acuna, E., Rodriguez, C., 2004. The treatment of missing values and its effect in the classifier accuracy. In: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), pp. 639–47.

ARPA-Puglia, 2017. Available at http://www.arpa.puglia.it/web/guest/serviziometeo (accessed on 10.11.2017).

Avidov, Z., 1954. Further investigations on the ecology of the olive fly (Dacus oleae Gmel.) in Israel. Ktavim 4, 39–50.

Barzman, M., Bàrberi, P., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Graf, B., Hommel, B., Jensen, J.E., Kiss, J., Kudsk, P., Lamichhane, J.R., Messéan, A., Moonen, A.C., Ratnadass, A., Ricci, P., Sarah, J.L., Sattin, M., 2015. Eight principles of integrated pest management. Agron. Sustainable Dev. 35 (4).

Bimonte, S., Boulil, K., Pinet, F., Kang, M.A., 2013. Design of complex spatio-multidimensional models with the ICSOLAP UML profile-an implementation in MagicDraw. In ICEIS (1), pp. 310–315.

Bimonte, S., Chanet, J.P., Capdeville, J., Lefrileux, Y., 2014. Energetic assessment of dairy activities using OLAP systems. In: Ageng 2014 International Conference on Agricultural Engineering, vol. 6, No. 07, pp. 2014–10.

Bimonte, S., Pradel, M., Boffety, D., Tailleur, A., André, G., Bzikha, R., Chanet, J.P., 2013b. A New sensor-based spatial OLAP architecture centered on an agricultural farm energy-use diagnosis tool. Int. J. Decision Support Syst. Technol. (IJDSST) 5 (4), 1–20.

Borchert, R., Rivera, G., 2001. Photoperiodic control of seasonal development and dormancy in tropical stem-succulent trees. Tree Physiol. 21 (4), 213–221.

Burrack, H.J., Zalom, F.G., 2008. Olive fruit fly (Diptera: Tephritidae) ovipositional preference and larval performance in several commercially important olive varieties in California. J. Econ. Entomol. 101 (3), 750–758.

Caffarra, A., Eccel, E., 2010. Increasing the robustness of phenological models for Vitis vinifera cv. Chardonnay. Int. J. Biometeorol. 54 (3), 255–267.

Chandler, D., Bailey, A.S., Tatchell, G.M., Davidson, G., Greaves, J., Grant, W.P., 2011. The development, regulation and use of biopesticides for integrated pest management. Philos. Trans. Royal Soc. London B: Biol. Sci. 366 (1573), 1987–1998.

Chaturvedi, K., Rai, A., Dubey, V., Malhotra, P., 2008. On-line analytical processing in agriculture using multidimensional cubes. J. Ind. Soc. Agril. Statist 62 (1), 56–64.

Chaudhary, S., Sorathia, V., Laliwala, Z., 2004. Architecture of sensor based agricultural information system for effective planning of farm activities. In: Services Computing, 2004. (SCC 2004). Proceedings. 2004 IEEE International Conference on. IEEE, pp. 93–100.

Childes, S.L., 1988. Phenology of nine common woody species in semi-arid, deciduous Kalahari Sand vegetation. Vegetatio 79 (3), 151–163.

Chuine, I., 2000. A unified model for budburst of trees. J. Theor. Biol. 207 (3), 337–347.

Crovetti, A., Quaglia, F., Loi, G., Rossi, E., Malfatti, P., Chesi, F., Conti, B., Belcari, A., Raspi, A., Paparatti, B., 1982. Influence of temperature and humidity on the development of the immature stages of Dacus oleae (Gmelin). Frustula Entomol. 5, 133–166.

Deggau, R., Fileto, R., Pereira, D., Merino, E., 2010. Interacting with spatial data warehouses through semantic descriptions. GeoInfo 2010, 122–133.

Donatelli, M., Bellocchi, G., Fontana, F., 2003. RadEst3. 00: software to estimate daily radiation data from commonly available meteorological variables. Eur. J. Agron. 18 (3), 363–367.

Donatelli, M., Campbell, G.S., 1998. A simple model to estimate global solar radiation. In: Proceedings of the Fifth Congress of the European Society for Agronomy. Nitra, Slovakia, II, pp. 133–134.

Easterling, W.E., Aggarwal, P.K., Batima, P., Brander, K.M., Erda, L., Howden, S.M., Kirilenko, A., Morton, J., Pingali, P., Soussana, J.F., Tubiello, F.N., 2007. Food, fibre and forest products. Clim. Change 273–313.

Parliament, European., 2009. Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides. Off. J. Eur. Union 309, 71–86.

FAOSTAT (2014) available at http://www.fao.org/faostat/en/#data/QC (accessed on 20. 11.2017).

Fletcher, B.S., Pappas, S., Kapatos, E., 1978. Changes in the ovaries of olive flies (Dacus oleae (Gmelin)) during the summer, and their relationship to temperature, humidity and fruit availability. Ecol. Entomol. 3 (2), 99–107.

Fountas, S., Carli, G., Sørensen, C.G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., Liakos, B., Canavari, M., Wiebensohn, J., Tisserye, B., 2015. Farm management information systems: Current situation and future perspectives. Comput. Electron. Agric. 115, 40–50.

Gallo, C., De Bonis, M., Perilli, M., 2010. Data warehouse design and management: theory and practice. IEEE Members. Dipartimento di Scienze Economiche, Matematiche e Statistiche Università Di Foggia Largo Papa Giovanni Paolo II, 1–71121.

García de Cortázar-Atauri, I., Brisson, N., Seguin, B., Gaudillere, J.P., Baculat, B., 2005. Simulation of budbreak date for vine. The BRIN model. Some applications in climate change study. In: XIV International GESCO Viticulture Congress, Geisenheim, Germany, 23–27 August, 2005. Groupe d'Etude des Systemes de COnduite de la vigne (GESCO), pp. 485–490.

Genç, H., Nation, J.L., 2008. Survival and development of Bactrocera oleae Gmelin (Diptera: Tephritidae) immature stages at four temperatures in the laboratory. Afr. J. Biotechnol. 7 (14).

Granier, C., Tardieu, F., 1999. Water deficit and spatial pattern of leaf development. Variability in responses can be simulated using a simple model of leaf development. Plant Physiol. 119 (2), 609–620.

Gregory, P.J., Ingram, J.S., Campbell, B., Goudriaan, J., Hunt, L.A., Landsberg, J.J., Linder, S., Stafford Smith, M., Sutherst, R.W., Valentin, C., 1999. Managed production systems.

Gupta, A.K., Mazumdar, B.D., 2013. Multidimensional schema for agricultural data warehouse. Int. J. Res. Eng. Technol. 2 (3), 245–253.

Häkkinen, R., Linkosalo, T., Hari, P., 1998. Effects of dormancy and environmental factors on timing of bud burst in Betula pendula. Tree Physiol. 18 (10), 707–712.

Hay, R.K.M., Porter, J.R., 2006. The Physiology of Crop Yield, second ed. Blackwell Publishing, Oxford UK.

Hegazi, E., Herz, A., Hassan, S.A., Khafagi, W.E., Agamy, E., Zaitun, A., et al., 2007. Field efficiency of indigenous egg parasitoids (Hymenoptera, Trichogrammatidae) to control the olive moth (Prays oleae, Lepidoptera, Yponomeutidae) and the jasmine moth (Palpita unionalis, Lepidoptera, Pyralidae) in an olive plantation in Egypt. Biol. Control 43 (2), 171–187.

Italian Institute of Statics, 2010. Agricultural Census. Available at: http://dati-censimentoagricoltura.istat.it/ accessed 20.03.17.

Italian Institute of Statics, 2016. Area and Production. Available at: www.istat.it/it/ accessed 20.03.17.

Jolly, W.M., Nemani, R., Running, S.W., 2005. A generalized, bioclimatic index to predict foliar phenology in response to climate. Glob. Change Biol. 11 (4), 619–632.

Kimball, R., Ross, M., 2013. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.

Kramer, K., 1996. Phenology and growth of European trees in relation to climate change. Kramer.

Larcher, W., & Bauer, H.L., 1981. Ecological significance of resistance to low temperature. In: Physiological Plant Ecology I. Springer, Berlin Heidelberg, pp. 403–437.

Lechowicz, M.J., Koike, T., 1995. Phenology and seasonality of woody plants: an unappreciated element in global change research? Can. J. Bot. 73 (2), 147–148.

Levitt, J., 1980. Responses of Plants to Environmental Stress, Volume 1: Chilling,

Freezing, and High Temperature Stresses. Academic Press.
Lieberman, D., 1982. Seasonality and phenology in a dry tropical forest in Ghana. J. Ecol. 791–806.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 50–60.
McFadden, M.W., Kapatos, E., Pappas, S., Carvounis, G., 1977. Ecological studies on the olive fly Dacus oleae Gmel. in Corfu. I. The yearly life cycle. Bollettino.
Miquel, M., Bédard, Y., Brisebois, A., Pouliot, J., Marchand, P., Brodeur, J., 2002. Modeling multi-dimensional spatio-temporal data werehouses in a context of evolving specifications. Int. Arch. Photogrammetry Remote Sens. Spatial Information Sci. 34 (4), 142–147.
Miranda, C., Santesteban, L.G., Royo, J.B., 2013. Evaluation and fitting of models for determining peach phenological stages at a regional scale. Agric. For. Meteorol. 178, 129–139.
Monteith, J.L., Unsworth, M.H., 1990. Principles of Environmental Physics. Edward Arnold, New York, NY.
Mott, K.A., Parkhurst, D.F., 1991. Stomatal responses to humidity in air and helox. Plant, Cell Environ. 14 (5), 509–515.
Nilakanta, S., Scheibe, K., Rai, A., 2008. Dimensional issues in agricultural data warehouse designs. Comput. Electron. Agric. 60 (2), 263–278.
Njoku, E., 1958. The photoperiodic response of some Nigerian plants. J. West African Sci. Assoc. 4, 99–111.
Orlandi, F., Garcia-Mozo, H., Dhiab, A.B., Galán, C., Msallem, M., Romano, B., Abichou, M., Dominguez-Vilches, E., Fornaciari, M., 2013. Climatic indices in the interpretation of the phenological phases of the olive in mediterranean areas during its biological cycle. Clim. Change 116 (2), 263–284.
Osonubi, O., Davies, W.J., 1980. The influence of plant water stress on stomatal control of gas exchange at different levels of atmospheric humidity. Oecologia 46 (1), 1–6.
Parker, A.K., DE CORTÁZAR-ATAURI, I.G., Van Leeuwen, C., Chuine, I., 2011. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. Aust. J. Grape Wine Res. 17 (2), 206–216.
Partanen, J., Koski, V., Hänninen, H., 1998. Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (Picea abies). Tree Physiol. 18 (12), 811–816.
Pestana, G., da Silva, M.M., Bédard, Y., 2005. Spatial OLAP modeling: an overview base on spatial objects changing over time. In: Computational Cybernetics, 2005. ICCC 2005. IEEE 3rd International Conference on. IEEE, pp. 149–154.
Pinet, F., Schneider, M., 2010. Precise design of environmental data warehouses. Oper. Res. Int. J. 10 (3), 349–369.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100 (2), 81–92.
Prokopy, R.J., 2003. Two decades of bottom-up, ecologically based pest management in a small commercial apple orchard in Massachusetts. Agric. Ecosyst. Environ. 94 (3), 299–309.
Rea, R., Eccel, E., 2006. Phenological models for blooming of apple in a mountainous region. Int. J. Biometeorol. 51 (1), 1–16.
Romero, O., Abelló, A., 2009. A survey of multidimensional modeling methodologies. Int. J. Data Warehouse. Min. 5 (2), 1.
Rosenthal, S.I., Camm, E.L., 1997. Photosynthetic decline and pigment loss during autumn foliar senescence in western larch (Larix occidentalis). Tree Physiol. 17 (12), 767–775.
Salah, H.B.H., Tardieu, F., 1996. Quantitative analysis of the combined effects of temperature, evaporative demand and light on leaf elongation rate in well-watered field and laboratory-grown maize plants. J. Exp. Bot. 47 (11), 1689–1698.
Sautot, L., Bimonte, S., Journaux, L., Faivre, B., 2015. Dimension enrichment with factual data during the design of multidimensional models: application to bird biodiversity. In: International Conference on Enterprise Information Systems. Springer, Cham, pp. 280–299.
Schwartz, M.D., 1999. Advancing to full bloom: planning phenological research for the 21st century. Int. J. Biometeorol. 42 (3), 113–118.
Tenhunen, J.D., Lange, O.L., Jahner, D., 1982. The control by atmospheric factors and water stress of midday stomatal closure in Arbutus unedo growing in a natural macchia. Oecologia 55 (2), 165–169.
Tsitsipis, J.A., 1977. Effect of constant temperature on the eggs of the olive fruit fly, Dacus oleae (Diptera, Tephritidae). In Annales de Zoologie Ecologie Animale.
Tzanakakis, M.E., 2006. Insects and Mites Feeding on Olive: Distribution, Importance, Habits, Seasonal Development and Dormancy. Brill Academic Publishers.
Wang, X.G., Johnson, M.W., Daane, K.M., Opp, S., 2009. Combined effects of heat stress and food supply on flight performance of olive fruit fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 102 (4), 727–734.
White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. Global Biogeochem. Cycles 11 (2), 217–234.
Xu, H., Twine, T.E., Yang, X., 2014. Evaluating remotely sensed phenological metrics in a dynamic ecosystem model. Remote Sens. 6 (6), 4660–4686.
Zaza, C., Bimonte, S., Faccilongo, N., La Sala, P., Contò, F., Gallo, C., 2018. BI4IPM: a business intelligence system for the analysis of olive tree's integrated pest management. Int. J. Agric. Environ. Inf. Syst. (IJAEIS) 9 (1), 16–38.
Zimmermann, M.H., 1964. The effect of low temperature on ascent of sap in trees. Plant Physiol. 39 (4), 568.