An Integrated Framework for Predicting Long-Term Productivity of Pastures in the Kingdom of Saudi Arabia

Ahsan Abdullah¹ · Ahmed Bakhawain² · Abdullah Basuhail¹ · Ahtisham Aslam¹

Received: 5 March 2015 / Accepted: 10 August 2015 / Published online: 14 September 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract The population of Saudi Arabia is increasing so is the demand for food; however, the arable land that can support this demand is decreasing rapidly. To meet the increasing dietary (cereal, meat, milk, etc.), needs of people and the fodder needs of livestock require identification of additional cultivation regions and correspondingly suitable crop/grass varieties. The traditional methods to achieve these objectives are expensive, complex and time-consuming. Therefore, the exploration of novel and proven IT techniques and methodologies are needed to address this complex problem. In this paper, we propose a data-driven framework and present simulated results mapped to real data that show how predictive data mining, geographical information system and expert system can be integrated. This integration results in identifying promising cultivable regions for the long-term productivity of perennial pasture grasses in the Kingdom of Saudi Arabia. The proposed framework can ultimately assist in identification of promising rangeland areas, the identified areas subsequently explored as per necessary follow-up actions/procedures.

Keywords Data mining · Image processing · Geographical information system (GIS) · Perennial grass · Agriculture · Meteorology · Ecosystem · Environment

1 Introduction

The population of the Kingdom of Saudi Arabia (KSA) can be characterized by rapid growth; from Fig. 1, it can be observed that over the last decade, there has been a sharp increase in the Kingdom's population. The increase in population results in an increased demand for food (cereal, meat, milk, eggs, etc.); this increase can either be met locally by increasing yield/production or the alternate option of food imports. The meat consumption ecosystem is, however, complex, which not only consists of population demographics, but also economic growth, crop yields, animal yields, arable land, agrometeorological parameters, climate change to name a few and is discussed further.

An important parameter of that ecosystem is cereal yield. Cereal yield [1] is measured as kilograms per hectare (kg/ha) of harvested land, which includes wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat and mixed grains. Other than specialized fodder crops such as alfalfa and Rhodes grass by-products of cereal crops are also used as fodder and feed for livestock. As per World Bank [2], the cereal yield of the Kingdom is on a decline, the latest figures being 4120 kg/ha.

Another important parameter of that ecosystem is beef yield. The average beef yield (boneless steaks and roasts) from the current breed of a 1200-pound live cow is approximately 425 lbs [3] or 193 kg. As per Guardian [4], it takes 7 kg of feed to produce 1 kg of beef. Thus, currently 1 ha in the Kingdom can only yield fodder to produce beef equivalent to three cows. Despite the ups and downs in the economic growth, during the last decade, there has been a sharp increase in beef consumption in KSA, which can be observed from Fig. 2 [5].

However, review of the arable land that can be used to produce feed and fodder to raise livestock in the Kingdom

✉ Ahsan Abdullah
ahsan1010@yahoo.com; aabdullah1@kau.edu.sa

¹ Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

² Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

Fig. 1 Increase in population of KSA over the last decade.

Source: World Bank

Fig. 2 Beef consumption in KSA over the last two decades [5]

points to an alarming situation, i.e., a sharp decline; arable land has actually decreased by 10 % over the last decade as can be seen from Fig. 3.

This is not the end, in view of Fig. 2, the demand for meat is expected to continue to rise as the prosperous population is increasing and will ultimately reach a saturation point. A rising middle class desiring more animal source foods and dairy products resulting in an increase in demand for fresh water and arable land, which in the case of Saudi Arabia, is on a decline [6]. Thus, there is a critical need to predict and identify potential cultivable regions in the Kingdom that are suitable for further study and exploration.

Fig. 3 Decrease in arable land (% of land area) of KSA during the last decade. Ref: World Bank

In addition to this, extreme agrometeorological parameters, e.g., surface temperature (which is 41–48 °C in the summer), annual rainfall (which is only 100–150 mm) and high solar radiation are among the hurdles in locally increasing the yield. In view of these, agrometeorological parameters applying traditional approaches such as engaging in livestock production for a period of only 3–4 months over the year-round grazing of the rangeland are not enough to meet the fodder needs of the Kingdom. Therefore, to address the challenges to the vulnerable ecosystem effected by population dynamics, decrease in arable land, decrease in yield and the difficulties associated with cultivation due to the extreme climatic factors, there is a need to develop a decision support system (DSS). A DSS can process large multi-attribute data sets so as to identify (i) promising uncultivated regions in the Kingdom and (ii) also identify those varieties of pasture grass that are suitable for cultivation in the identified regions.

1.1 Prior Solutions

Traditional predictive modeling methods, such as additive (e.g., logistic regression) or generalized linear, are finite-dimensional models that are based on apriority assumptions on the shape of the response of (say) fodder to (say) agrometeorological factors. This approach may be rather naïve, as fodder species often show mixed and complex response to

variations of agrometeorological parameters [7]. Therefore, higher-order interaction terms are required to be included to deal with skewed or multimodal response shapes; however, this often leads to false and biologically impractical solutions that are ecologically difficult to interpret [8].

To avoid these issues, a number of modern nonparametric multiplicative regression techniques have been developed and reported in the literature that do not make any apriority assumptions on the shape of fodder species responses (in the context of our work) to agrometeorological predictor parameters [9]. Some multi-purpose nonparametric models, such as support vector machines (SVMs), artificial neural networks (ANNs), multivariate adaptive regression splines (MARS) or maximum entropy algorithms (MAX-ENT) have been reported to be successfully applied to a broad range of prediction problems [10], including biogeographical and ecological questions such as fodder species distributions.

More specifically, different research contributions, such as [15], have adopted an approach of using one or two statistical or computational techniques for addressing agriculture and its related problems. In [16], the authors presented the use of FAUSY (fuzzy algorithm) to extract relationships from the set of input–output environmental observations and then applied the machine-learning technique for decision support. Keeping in mind the rapidly growing role of developing decision support system and expert system (ES) to solve domain-specific problems, in [17, 18], the authors propose DSS and ES, respectively, that can be used for the prediction of climatic factors and their effects on the crop production. In [19], the authors present an integrated solution of information and communication technologies (ICT) and DSS with web-based interface and mobile connectivity to share live data. In that solution, a rule-based approach is used for predicting the results based on different climatic factors. In [20], the author presents an agriculture DSS that helps small and large landowners with their decisions by identifying the relationship between Bt and non-Bt cotton cultivation based on primary, raw-data analysis. Parallel to all of this, in [12, 14, 22], the authors describe the role of using information technology (IT) in agriculture to solve the related problems.

We believe that the solutions and methods [11–21] have some limitations, e.g., they are expensive, time-consuming to learn, complex and not efficient for decision support applications. Furthermore, these methods/solutions are not able to answer critical questions, such as the relationship of arable land with respect to inland features (such as vegetation, population centers and highways) in arable patches and considering their impact on production, while using the geometric network model. According to Wang et al. [22], the arable land identification problem has been studied by taking into consideration the inland factors; however, neither

the predictive data mining as proposed in our framework, nor spatial querying has been used (which we do) to take into consideration the inland features. Also, most of the existing solutions are designed and implemented for transactional processing and minimal analysis approach, while the problem under consideration needs a thorough and detailed predictive data-driven analysis, which indeed needs an integrated solution of proven techniques focusing on agriculture [1].

1.2 Contributions and Advantages of Our Work

- We have proposed an integrated framework to alleviate the problem of decline in arable land and increase in food demand for the Kingdom of Saudi Arabia. System design based on feedback of 130 local professionals. The proposed system is to identify cultivable regions that are currently not under cultivation.
- As per the literature review and to the best of our knowledge, no indigenous work as per the local needs and similar to ours has been undertaken for the Kingdom. Hardly any joint R&D cross-discipline collaborations have been reported to exist between IT and agriculture (range plant) domain experts of the region to address the real-life problem as addressed in our work.
- The traditional approach of identifying feasible range-land areas is time-consuming and also expensive. The proposed framework will allow decision makers, researchers, academicians, etc. to quickly process large data sets to identify and evaluate potential cropping/fodder regions, along with computed probabilities of success of the selected species of crop/fodder. This will help enhance the area under cultivation, increase yield, increase healthy livestock population, thus increasing economic prosperity of the Kingdom.
- Under the King Abdullah's Agricultural Initiative [23], the policy of direct investment in foreign agriculture is being pursued. The proposed spatial neutral framework is aligned with this policy and can be used to identify promising cultivable regions outside the Kingdom to help ensure strategic food security of the Kingdom.
- Tools like ArcGIS and QuantumGIS (QGIS) are excellent at processing, projecting spatial data, resampling raster layers, map preparation, etc. but do not have built-in predicative modeling capability as our framework-based tool does.
- Predictive modeling is built into the proposed tool; thus, the user is not required to do Linux shell scripting or R language programming for statistical computing/modeling or use AWK programming language for processing text-based data or Gnuplot program for 2D or 3D plots, etc.

1.3 Overview of Our Work

In this paper, we present a framework that endeavors to address the limitations of the existing prior solutions (Sect. 1.1) by providing an integrated approach consisting of proven information technology techniques of data mining, ESs, image processing, spatial data warehousing (SDW), and geographical information system (GIS). In our framework, we use ‘predictive’ data mining, i.e., we firstly ‘train’ our system by using the existing agrometeorological data and use a particular data mining technique (Naïve Bayesian classifier) so as to predict the viability of cultivating a particular fodder variety under particular agrometeorological parameters for identifying promising cultivable regions.

We address the problem by considering several biophysical factors (e.g., soil, altitude and temperature), artificial factors (e.g., roads and markets) as well as socioeconomic factors (e.g., population, culture and market value) as shown in Fig. 4. Note that all of these factors may make a crop infeasible for cultivation in a region that has been identified as ‘feasible’ based on agrometeorological factors and vice versa. Consider Fig. 5a that shows agriculture farms in the northwest of the Kingdom (IKONOS satellite 0.8 m resolu-

tion, Tadco Farms, Tabuk, Saudi Arabia), there are other such farms near the eastern coast of the Kingdom too, while the market for the produce may be elsewhere. For example, Fig. 5b shows fodder grown elsewhere being sold at the roadside market near Jeddah at the southeast of the Kingdom.

Our integrated framework is envisioned to counter these issues by adopting a data-driven approach of spatial queries. The challenges being, unavailability of georeferenced data, even if the data are available, data are diverse in terms of scale, resolution, scope, etc. and contains a great deal of ‘noise’ making its optimum use difficult. Therefore, image inpainting techniques [24,25] are used (which is part of the framework [26]) to remove ‘noise’ (such as city names, administrative boundaries, roads and other symbols/icons) from the available maps in digital or hard copy format. Once these maps are cleaned and the ‘noise’ is removed and replaced with the correct data, using our framework the raster maps are converted into comma separated values (CSVs) and then loaded into the spatial data warehouse (SDW) as georeferenced tables. These tables contain the Cartesian coordinate data as well as the RGB values for these coordinates stored in the SDW. The predictive data mining technique (of Naïve Bayesian classifier) is subsequently used that produces prob-

Fig. 4 Major factors affecting the success of cultivating a crop/grass in a given region

Fig. 5 Cultivation and sale of fodder in different regions of the Kingdom. **a** Farms in Tabuk (north of Kingdom), **b** roadside fodder market near Jeddah (southwest of Kingdom)

abilistic classes, i.e., when presented with unclassified data item, the Naïve Bayes model presents the probability that the new unclassified item belongs to which of the possible class categories [27]; these results are then color-coded and displayed in the GIS environment of the framework.

1.4 Naïve Bayesian Classifier

In this section, we will briefly describe the mathematical formulation of the Naïve Bayesian classification technique used in our proposed framework. Bayes' theorem is from the domain of probability theory originally stated by the Reverend Thomas Bayes. The theorem helps in understanding how the probability that a hypothesis is true is affected by newly presented evidence. The theorem has been used in a wide variety of contexts, ranging from development of 'Bayesian' spam blockers for email systems to marine biology and image recognition systems. With regard to the philosophy of science, the theorem has been used to try to elucidate the relationship between hypothesis and evidence. Bayes' theorem has also been used to provide insights into falsification, confirmation, the relation between science and pseudo-science, and making other topics more precise, corrected or sometimes extended. The following formulation of Bayes' theorem is as per [28]; details of our corresponding prediction module are given in Sect. 3.7.

Let A and B_j be sets. As per conditional probability

$$P(A \cap B_j) = P(A)P(B_j | A), \quad (1)$$

where $\cap$ denotes intersection (AND), and also that

$$P(A \cap B_j) = P(B_j \cap A) = P(B_j)P(A | B_j) \quad (2)$$

Therefore,

$$P(B_j | A) = \frac{P(B_j)P(A | B_j)}{P(A)} \quad (3)$$

Now, let

$$S \equiv \bigcup_{i=1}^N A_i, \quad (4)$$

so A_i is an event in S and $A_i \cap A_j = \phi$ for $i \neq j$, then

$$A = A \cap S = A \cap \left(\bigcup_{i=1}^N A_i \right) = \bigcup_{i=1}^N (A \cap A_i) \quad (5)$$

$$P(A) = P\left(\bigcup_{i=1}^N (A \cap A_i)\right) = \sum_{i=1}^N P(A \cap A_i). \quad (6)$$

However, this can be written as

$$P(A) = \sum_{i=1}^N P(A_i)P(A | A_i) \quad (7)$$

So

$$P(A_i | A) = \frac{P(A_i)P(A | A_i)}{\sum_{j=1}^N P(A_j)P(A | A_j)} \quad (8)$$

The remainder of the paper is organized as follows: Sect. 2 describes the related work from IT as well as agriculture and long-term perspectives. In Sect. 3, we outline the modular architecture of our proposed framework. Comparative study of different data mining techniques, extract transform load (ETL) and data validation w.r.t framework is discussed in Sect. 4. Simulated results are presented in Sect. 5, and finally, we conclude our work in Sect. 6.

2 Related Work

2.1 The IT Perspective

In this section, we will discuss some existing work on the application of IT and its related technologies to address agriculture and related problems.

In [16], the authors present an intelligent system that can be used to capture the relationship between environmental variables and sap-flow measurements. The system makes use of the 'fuzzy algorithm' (i.e., FAUSY) to extract relationships from a set of input–output environmental observations and applies algorithmic techniques for learning and forecasting via a simulation model. These predicted results help farmers and policy makers in planning and developing future cultivation programs.

A DSS proposed in [17] takes into account different factors such as weed control, pests, crop selection, soil factors (e.g., soil flora and soil fauna) and helps in decision making to handle conventional and genetically modified crops. Instead of going to the micro-level functions of individual species, soil fauna modeling was done at the community level by making use of machine-learning methods such as regression trees, model trees and linear equations. Finally, the proposed system was used for the prediction of climatic factors and their possible effects on crop production.

Agro Genius [18] is an ES that is designed to help different stakeholders at Informatory and Advisory Levels. At the Informatory Level, users can obtain static information about the advantages and disadvantages of organic and inorganic cultivation under certain weather conditions. At the Advisory Level, users can interact with the system in the form of a question–answer session, resulting in a flow chart-like

Table 1 The description of six grass cultivars [28]

English name	Scientific name	Cultivar name	Country of origin
Perennial grass	<i>Lolium perenne</i>	Aries	New Zealand
Perennial grass	<i>Lolium perenne</i>	Quarter	New Zealand
Orchardgrasses	<i>Dactylis glomerata</i>	Tekapo	New Zealand
Orchardgrasses	<i>Dactylis glomerata</i>	Niva	Czech Republic
Tall fescues	<i>Festuca arundinacea</i>	K5666v	New Zealand
Tall fescues	<i>Festuca arundinacea</i>	Fawn	Oregon-USA

tree structure. The generated ‘tree’ shows nodes, leaves, etc. as climatic effects, weather forecasts, possible crops along with the possibility of their successful growth under these climatic conditions.

In [19], a joint solution of DSS and ES is presented that can help small-holding farmers in decision making by minimizing the impact of climatic risk factors and assisting farmers by suggesting the best possible options for improving crop productivity. The solution [19] also enables the stakeholders to share this agricultural information among themselves (by using mobile phone connectivity) and to keep the system updated with the prevailing climatic conditions and weather factors. Taking into consideration the impact of different parameters, the DSS uses a rule-based approach for prediction and shares the prediction results and suggestions with the end users.

In addition to the already discussed related work, many other researchers (both from academia and from industry) have presented IT-based solutions to address agriculture-related problems (e.g., [12, 14, 20, 21]). Our initial studies indicate that in a region such as Saudi Arabia, with its particular climatic factors, weather effects and nature of soil changes, a single ES- or DSS-based approach may not be sufficient (as also mentioned in [1]). We, therefore, need a system which is a combination of ES, DSS and GIS and uses proven predictive data mining technique to address the discussed problem—our proposed framework is an endeavor in this direction.

2.2 The Agricultural Perspective

In this section, we focus on the related work from the perspective of agriculture, specifically investigating those pasture grass varieties that can provide maximum yield over the year. For example, Alfalfa (*Medicago sativa* L.) gives an excellent production of total dry matter yield (DMY), which was estimated to be two million tons in 2008 (Saudi Arabia Agriculture Statistical Yearbook, 2009). However, the increasing needs of livestock production and the excessive requirement of water by Alfalfa, the researchers are strongly urged to investigate alternate species [28]. At the same time, the extreme desert environment and heat stress are two major

Table 2 Possible fodder species from the Khulais region, KSA [29]

Species name	Life-form
<i>Cenchrus ciliaris</i> L.	Hemi-cryptophytes
<i>Cenchrus pennisetiformis</i> Hochst. & Steud.	Therophytes
<i>Eragrostis minor</i> Host.	Therophytes
<i>Panicum coloratum</i> L.	Hemi-cryptophytes
<i>Panicum turgidum</i> Forssk.	Hemi-cryptophytes
<i>Pennisetum setaceum</i> (Forssk.) Chiov.	Hemi-cryptophytes
<i>Pennisetum divisum</i> (J. F. Gmel.) Henrard	Hemi-cryptophytes
<i>Setaria verticillata</i> (L.) P. Beauv.	Therophytes
<i>Sorghum halepense</i> (L.) Pers.	Therophytes
<i>Stipa capensis</i> Thunb.	Therophytes
<i>Stipa parviflora</i> Desf.	Hemi-cryptophytes

factors that affect the production of forage in the Kingdom. The study [28] was carried out on six grass cultivars (i.e., two perennial ryegrasses, two endophyte-free tall fescues and two orchard grasses) (please see Table 1) to evaluate yield production under these extreme weather conditions [28].

Alsherif et al. [29] studied the flora of the Khulais region (west Saudi Arabia), with reference to the region’s potential use and identified 66 species that are used as fodder; some of those species that can be considered for further investigation for cultivation are listed in Table 2.

While investigating the result of local agrometeorological parameters on crop cultivation using traditional methods, we found some of those parameters to be promising for the cultivation of the six species (Table 1) and 11 out of the 66 selected fodder species (Table 2) [29]. However, when we consider the entire Kingdom with scores of such parameters, the complexity of the problem increases tremendously and is well beyond the manual methods. Previous studies have indicated that producers should select cultivars that originate from a location with similar climatic conditions [30]. Since the indicated species have adapted to a wide range of environmental conditions, once these cultivars are identified, the farmers in Saudi Arabia can benefit from growing them. However, there is no research available that makes a valid recommendation on how these species would perform under

the excessively high temperature that prevails in the central region of Saudi Arabia [28].

2.3 The Agrometeorological Perspective

There are various factors that affect the delicate meat consumption ecosystem including (but not limited to): climate, diseases and parasites [31]. Climate is the most significant factor where patterns of rainfall and temperature significantly impact the growth of fodder and pasture land that is used for cultivation throughout the year [32].

With the increasing demand for food, there is a need to predict that demand for timely mitigation action/policies; therefore, many researchers have addressed the corresponding issues using traditional means, for example [33–35]; however, these researchers have not used an integrated predictive data mining GIS approach as we have proposed in this paper. In [33–35], the rise in atmospheric concentration of CO₂ and subsequent climatic changes has been discussed for 2050. As per [33], climate change may also adversely affect the prospect of achieving food security, since most climate models indicate that agricultural potential of Middle East countries may be affected more than the world average. The high dependence of several countries of this region on food import makes them particularly vulnerable in this respect. Different [34,35] climatic factors will drive global warming; factors like CO₂ fertilization can also have a positive impact on tree and crop growth with increase in biomass production; unfortunately, this is not the case for countries of this region [36]. In [37], forecast of red beef is studied for various Gulf Cooperation Council (GCC) countries from 2012 to 2015. Population growth in the GCC countries is predicted to increase by 7 % with the total population reaching over 47 million in 2015. By 2017, food consumption is forecasted to increase in Saudi Arabia by 53.3 %, but at the same time shrinking arable land to grow food and fodder points to an alarming situation. Thus, there is a need to predict and identify potential cultivable regions in the Kingdom for early mitigation and necessary actions.

2.4 The Long-Term Perspective

One categorization of plants can be based on the duration of their life cycles, i.e., annual versus perennial. Annual plants last a year in their native environment and ecosystem, while perennials live additional years as compared to the annual plants. Figure 6 shows two of the many perennial grasses of the Kingdom, i.e., Alfalfa and Blue Panc sampled during our *Hada Al-Sham* farm field visit. Recognize that some plants that are perennials in their native (e.g., tropical) habitat and ecosystem are treated as if they were not perennial in colder regions. For example, lantana plants are, technically, perennial, but they are considered as annual plants in northern regions, i.e., regions too cold for them to complete their natural life cycle [38].

Crop and fodder production consists of the right selection of annual and perennial crops, their cultivars and varieties. The objective is to meet local and market needs according to their site suitability and their role within the crop rotation and ecosystem. These critical decisions are made by taking into consideration the management of soil fertility, water, land use, climate, response to available inputs to name a few. The objective of the proposed framework discussed in this paper is to facilitate making such decisions by making them data-driven and using predictive modeling to take into consideration the long-term perspective, i.e., perennial crops, more specifically perennial grasses.

Other than perennial aspects of grasses, there are also other aspects of long-term productivity. For example, in the context of water, for long-term, some farmland may need to be removed from production or converted to other uses. Other uses include conversion of raw crop land for production of drought-tolerant forages. In the context of soil management, soil must be protected and nurtured to ensure its long-term productivity and stability. Methods to enhance and protect the yield of soil include reducing tillage, using cover crops, fertilizer and/or manures and maintaining soil cover. In the context of land use, conversion of agricultural land to urban uses as rapid growth and escalating land values threaten farming on

Fig. 6 Two of the perennial grasses of the Kingdom of Saudi Arabia being considered in our study. **a** Alfalfa, **b** Blue Panc

prime soils. However, in the context of this paper, we will restrict the scope to long-term aspects of perennial grasses only.

3 The Architecture of the Proposed Framework

Prior to the framework design, around 130 professionals from the agriculture and related domains from 36 organizations across the Kingdom were interviewed and surveyed using questionnaire and their feedback incorporated in the framework design. Figure 6 shows the breakdown of the education and related experience of the survey respondents. From Fig. 7a, it can be observed that although the respondents mostly held bachelor's degree, but about quarter of respondents were PhD and their feedback was highly valued.

Since the framework involves using digital data, GIS and forecasting systems and their integration, it was important for us to be aware of the related experiences and background of the respondents. From Fig. 7b, it can be observed that the usage and non-usage of cultivated crop yield maps are almost the same among the survey respondents; however, the use of

forecasting software and GIS-based applications was relatively high. Thus, we had the right mix of the respondents whose input was incorporated in the developing the framework; further discussion of the respondents and their survey is beyond the scope of this paper.

Figure 8 shows the implementation-oriented modular architecture of our proposed framework which consists of nine modules/components. Each of these modules/components is responsible for performing a specific task (as discussed in the subsequent subsections).

3.1 Graphical User Interface (GUI)

GUI is the first module of the framework from where users may initiate their interaction with the system. The GUI can be used to perform multiple tasks, e.g., to load the homogeneous data, perform spatial queries on available spatial data, perform 'mining' using the available data, view georeferenced results in the form of image overlays, loading and saving images and zooming in/out. The GUI of the system is proposed to provide a very simple means of interacting with the system so that, in addition to technical persons, non-

Fig. 7 Breakdown of the education and related experience of the survey respondents. **a** Breakdown of education, **b** breakdown of related experience

Fig. 8 Architecture of the proposed integrated framework consisting of nine modules

technical domain experts can also easily use it. The GUI can be employed by the end user to view the prediction results of their queries with color-coded probability of success of growing a particular variety of grass at identified locations. This module is to follow the GUI standards as per the visual and interactive guidelines of Microsoft.¹

3.2 The User Input Processor

This module is responsible for handling different user interactions along with passing control information so that the data can be routed and used by the textual data entry module or the image parser module. Suppose the user interacts with the system to load an image; this module distinguishes and understands the input command, i.e., either the input is a scanned image, textual data, spatial query, etc., and passes on the scanned image to the image parser and data extractor module for further processing and data extraction. Similarly, if the user gives a command via the GUI to perform a spatial query, this module receives the user input and pass on this query to the SDW for execution. We can say that this module plays the role of a ‘interface module’ between user inputs and other modules of the proposed framework.

3.3 The Image Parser and Data Extractor

This module is responsible for parsing raster maps in. TIF, BMP or other image formats and for removing ‘noise,’ i.e., unnecessary information from maps such as city names, administrative boundaries, roads and other symbols/icons. Although there are excellent commercial tools available for image processing, however, removing noise is easier said than done and replacing noise with credible data is a challenging task. Once the image is cleansed, i.e., clean image extracted from noisy image, the image is transformed into the required format (TIF, BMP, txt) and subsequently loaded, i.e., the ETL process performed. ETL issues are discussed in Sect. 4.

3.4 Textual Data Entry

Sometimes, tabular data may be required that is available in printed form, e.g., data that is published in books, research papers, reports or data could be available as text files or database files on CDs or magnetic tapes or available as web pages. Users can manually extract the required data from such printed sources, i.e., through data entry and, after the necessary transformation, can load the data into the SDW, i.e., ETL at the non-raster level. This module can also process and store lines and polygons that are either automatically converted or traced by a pen-tablet combination as SHP files.

Some of this work can also be done using commercially available tools such as VeryPDF that can convert a scanned PDF document into an editable Excel document (for more information, readers are referred to [39]). Then, there are tools like GSYS (www.jcprg.org/gsys/ver24/gsys-e.html) that can convert curve plots of scanned graph images to tables.

3.5 The Spatial Data Warehouse (SDW)

The SDW is designed using special schemas so as to provide high performance for analytical, low-selectivity queries that run on large data sets. The SDW is used to store data extracted from different data sources that has been cleaned by the ‘image parser and data extractor module’ (as discussed in Sect. 3.3). Once the clean data are loaded, the SDW module georeferences the image by using the key points (for georeferencing) that are provided by the user. This module can also be employed by the user to assign color codes and necessary labels to various parameter values. This module is also responsible for loading the legends of the raster maps, generating the corresponding RGB values and storing everything in the SDW. The framework envisages converting the image into a CSV text file which contains *x–y* coordinates, corresponding longitude–latitude values, legend values and RGB values for each pixel in the image. The longitude–latitude value is used as the primary key for each parameter value such that even if there is some missing data across a pair of parameters to be joined, the mismatching longitudinal–latitudinal values can be used for identification of missing values.

The outline of proposed process of raster to database conversion is shown in Fig. 9; here, PK is the primary key, which is subsequently converted to longitude and latitude.

Since the data generated by the framework are in CSV format, therefore, can be used as input by a number of popular data mining tools such as *Weka* (www.cs.waikato.ac.nz/ml/weka/) and *Orange* (www.orange.biolab.si), these are component-based data mining and machine-learning software suites that come with multiple classification and regression algorithms. Thus, the CSV format increases the scope of using the data generated by a variety of data mining tools, resulting in an increase in the scope of analysis.

3.6 Spatial Queries

Mainly, three operations related to SDW query processing are considered [40]: (i) joining large fact tables with large spatial/non-spatial dimension tables; (ii) computing one or more costly spatial predicates based on spatial ad hoc query windows; and (iii) aggregating data according to different levels of spatial granularity. For example, “Find the total area suitable for cultivation of a certain pasture grass variety inside a ‘rectangular’ window.” This query uses a topologi-

¹ <http://msdn.microsoft.com/en-us/library/aa511258.aspx>.

Fig. 9 The proposed process of converting raster maps into spatial database. **a** Raster map with legend, **b** legend text with RGB, **c** transformation from RGB to legend, **d** CSV file, **e** DB file

cal relationship and a spatial ad hoc query window that was not previously stored in the dimension tables. Another query can be used to ‘roll up’ to the granularity level of ‘city’ by using a larger window that identifies the cities where live-stock markets are located.

Other than these queries, some additional custom spatial queries are proposed as follows:

- Which markets can be served by cultivation in the predicted regions? Where can workers settle themselves for developing pastures in the predicted regions?
- Where are the population centers and roads w.r.t. the predicted regions? How far are they from the predicted regions?
- Buffering and display queries such as (i) displaying key points on the map such as cities and landmarks, (ii) buffering the geometry of an object (line, polygon) to a selected range in km, (iii) buffering a circle from the center of a selected polygon(s) on the map to a specified range in km, (iv) displaying the intersection of roads/wadis (valleys) with a selected polygon(s) on the map, and (v) identifying the overlap between specified regions and buffer from a particular city to a selected range in km. Visual results of some of the proposed queries are shown in Fig. 10.

3.7 The Predictive Data Mining Module

One of the important components of the proposed framework is the application of the predictive data mining technique that is appropriate for the data types considered in this framework. The Bayesian-based predictive data mining module takes user-selected factors/parameters from the GUI and then retrieves the historical data associated with these factors/parameters from the spatial database. Subsequently,

using the parameters selected by the user calculates the posterior probability of success of a selected attribute (e.g., grass variety) for each pixel. These probability results are then passed on to the output processor (Module-9), which assigns color codes to the probabilities via a color frequency distribution that corresponds to the calculated posterior probability. This is done in conjunction with the input of the user who defines probability classes such as ‘high,’ ‘medium’ and ‘low’ by selecting a range of probability values. Subsequently, color-coded, posterior-probability results are displayed on the map along with the corresponding legend.

3.8 The Overlay Formation Module

Overlay analysis is one of the spatial GIS operations that are used to observe and analyze the collective effect of different agrometeorological parameters such as temperature, % humidity, elevation and rainfall and their relationship in the overall ecosystem. Overlay analysis integrates spatial data with attribute data (attributes are the bits of information about each map feature) and makes it easy to understand the underlying ecosystem effecting the environment. Overlay analysis does this by combining information from one GIS parameter layer (for example, soil type) with another GIS layer (for example, temperature) to derive or infer the combined effect of the selected layers and how they affect each other.

3.9 The Output Processor

This module takes the results of Modules 6–8 and then displays them in a georeferenced environment and sends the combined result to the GUI module for display. This module can also convert the results into a standard vector file format such as SHP. Observe that the user commands (shown by red color) in Fig. 8 are proposed allow power users to access the

Fig. 10 Analogy visual results of spatial queries in the form of histograms

SDW through a command shell, while the feedback/replies (shown by blue color) in Fig. 8 are the responses to the commands executed by the power user.

4 Discussion

4.1 Data Mining Techniques

Prior to developing the prediction module, various predictive data mining techniques such as decision trees, SVMs, regression, C45, J48 and C5 were studied and evaluated in the context of agricultural forecasting and prediction. The objective being to determine the most appropriate technique that is expected to give the desired results. In our comparative study, we found some limitations in most of the above-mentioned techniques. For example, the reliability of decision trees is dependent on the quality of input data. Even a small change or ‘error’ in the input data can cause large variations in the results. Another major limitation of this technique is that the decisions/results are based on expectation and a small change in a rational expectation can lead to the very different results [41]. SVMs cannot perform well when the number of fea-

tures are greater than the number of samples. Also, another limitation is that SVM provide direct probability and we need to use additional cross-validation methods [42]. Regression could not be considered as a suitable data mining technique because it only looks at linear relationships between dependent and independent variables, i.e., it assumes a straight line relationship between variables, which is not usually true and may lead to unrealistic results [43]. C45 rules are slow for ‘noisy’ and large data sets [44]. Similarly, the major shortcoming of J48 is its run-time complexity which increases with the increase in depth of the ‘tree’ [45].

We found Naïve Bayesian modeling to be a better prediction technique, the reason being the technique provides specialized unsupervised learning and is least affected by the issues associated with other predictive data mining techniques. Naïve Bayesian modeling strengthens the model to perform advanced and detailed analyses such as the identification of pasture-cropping regions that require salinity control programs, etc. At the same time, Bayesian methods provide ‘formalism for reasoning under conditions of uncertainty, with degrees of belief coded as numerical parameters, which are then combined according to the rules of probability theory’ [46].

Naïve Bayesian modeling treats each parameter as an independent variable and keeps record of corresponding conditional probability of whether or not grass *G* grows in a particular region or not. In the ‘training’ phase, the proposed system is provided knowledge regarding the behavior of grass and then predict with certain probability whether or not grass *G* will grow in the regions where it is currently not cultivated. A properly trained system can determine the probability of success of cultivating grass species in an area for which agrometeorological parameter data are available.

4.2 ETL (Extract Transform and Load)

4.2.1 Noise in Raster Data

In the last section, we mentioned noisy and dirty data requiring cleansing, so what exactly is dirty data? There can be different types of noise; some of the noise types that we have considered in the framework are as follows:

- Inaccuracies introduced at the time of scanning
- Moiré patterns in scanned maps
- Errors due to vectoring
- Text and icons on the map
- Noise in satellite imagery
- Sliver, i.e., creation of additional polygons between adjacent polygons
- Polar to Cartesian coordinate transformational noise.

Some of the potential errors resulting in noise at the time of scanning are as follows:

- The accuracy or inaccuracy of the original drawing/map being scanned.
- The state of the map/drawing being scanned.

- The accuracy of the scanner itself. Large format or drum scanners are susceptible to inaccuracies if not calibrated frequently.
- The skill and care with which the map or part of it is scanned.
- The threshold settings of scanner.

Scanning of printed maps results in moiré patterns. A moiré pattern is an artifact or noise that is generated when two or more repeating patterns overlap. The most common cause of a moiré pattern in an image is when a printed image is scanned without first defocusing the image (Fig. 11a); however, defocusing results in loss of information. Moiré patterns are also caused by the frequency/angle of the scanner sensor (flat bed and drum scanner) harmonically repeating with a pattern in the object being scanned. Thus, the scanned image ends up having the moiré embedded as part of the image.

As compared to raster, vectors are more easily scaled, plotted and rotated; furthermore, they also require less storage space. If raster images could be consistently transformed into vector data, many operations could be performed efficiently on vectors. However, the process of vectorizing raster maps is subject to major uncertainties; thus, an infinite family of vector maps corresponds to each raster map [47]. Furthermore, the sheer number of single cell polygons that may exist in a satellite image may become a bottleneck during a raster-to-vector translation.

‘Noise,’ like city names, administrative boundaries, road networks, etc. (Fig. 7b), need to be removed, i.e., inpainting to get correct value of agrometeorological parameters against each pixel. However, removing the city names and/or administrative boundaries also ‘removes’ the data points that were ‘under’ the corresponding pixels.

The hypothetical latitude and longitude lines result in slight curves when drawn on flat maps; therefore, screen pixels being discrete points cannot capture the continuous

Fig. 11 Some sample cases of real ‘noise’ problems in raster data. **a** The moiré problem, **b** cluttering of text and icons, **c** slivers in overlaid images

nature of the curves, which can lead to errors. For example, in a typical 1:5,000,000 scale map Greenland appears to be of the same size as of the continent of Africa; however in reality, Africa is 14 times as large.

Slivers are created when polygon features share a boundary and a space between the features along the boundary. For example, there could be a gap due to an overlap between a coastal water area and the foreshore due to an island as shown in Fig. 11c. In both of these instances, the gap or overlapping area could be considered as a sliver polygon.

Noise in satellite imagery such as random variation impulsive noise, salt and pepper noise and speckle noise can be removed from satellite imagery using filters such as Wiener and Gaussian.

4.3 Validation of Results

Different statistical validation methods can be used to validate the framework results. One of these methods, i.e., k-fold validation which has some shortcomings, for example, we have to run the training algorithm k-times, which means it will take ‘k-times’ more computational resources and time. This means that k-fold validation cannot be used in scenarios where we have huge amounts of data [48]. For example, in our case, we have 100+ images/maps and most of these images are each more than 50 MBs in size (sometimes even exceeding more than 200 MBs in size for a single image). Processing these images results in 100 GBs of data, which is proposed to be stored in a high-performance Oracle SDW and used in the prediction process and running spatial queries. Keeping these points in mind, we use an alternate approach, i.e., we will randomly divide the data into ‘test’ and ‘training’ data sets for different k-times. This gives us the advantage to independently choose the size of the test set as well as the number of times that we repeat it. As part of this framework, we also propose validation at two more levels, i.e., overlaying results

on latest satellite imagery using *Google Earth* followed by ground-truthing, i.e., physically visiting the site to identify if the vegetation ‘seen’ to be cultivated via *Google Earth* is wheat or alfalfa or something else?

4.4 Overlay Analysis

In its simplest form, overlay analysis can be a visual operation, but analytical operations require one or more data layers to be joined physically by performing mathematical operations such as addition and subtraction. This overlay, or spatial join, can integrate data of different types such as soil, vegetation, land ownership and jurisdictions with the assessor’s parcels. The results of overlay analysis depend upon the spatial accuracy of the GIS layers and the quality of the GIS data being used. Weights can be assigned to maps as per order of placement in the map layer, and the transparency of maps can be changed by using transparency features. At a particular pixel, the combined effect (represented by a specific, unique color) of different parameters (actually represented by different colors in different layers, layer one, i.e., layer_1 through n th layer, i.e., layer_ n) can be calculated as:

$$\text{Pixel color} = (2^n * \text{layer}_1 + 2^{n-1} * \text{layer}_2 + \dots 2^1 * \text{layer}_n) * \text{CLR} / (2^n + 2^{n-1} + \dots 2^1)$$

Here, CLR is a unique identifier assigned to each color that is used to specify a color using RGB-based color model.

5 Results

5.1 Prediction Results and Validation

Using proven data mining techniques, the proposed system is initially ‘trained’ to use a subset of agrometeorological

Fig. 12 Color-coded probability of success of promising pasture grass cultivation regions. **a** Simulated prediction results, **b** prediction results overlaid on latest satellite imagery using *Google Earth*

Fig. 13 Overlay created by overlaying daily mean solar radiation map with agricultural lands map. **a** Daily mean solar radiation autumn, **b** agricultural lands, **c** both georeferenced clean maps overlaid

parameters such as rainfall and temperature. Based on these parameters, the proposed system will identify suitable fodder cultivation regions shown by a color-coded probability of success, the simulated results of which are shown in Fig. 12a. To verify the results, the current pasture-cropping regions are proposed to be overlaid with the predicted results as shown in Fig. 12a. Dividing the data into training and test sets and then comparing the results of the test set with the actual data will confirm the accuracy of the prediction results. Another approach is using (say decade old) historical data that does not show cultivation and then making predictions using that data and comparing results with latest satellite imagery. The analysis can be further enhanced by overlaying administrative boundaries, city names, roads, etc. on the prediction maps to help identify viable regions by inspection that is based on vicinities (better option being running spatial queries) to road networks, markets, population centers, etc. as shown in Fig. 12b which is created by overlying prediction results using Google Earth. The prediction module was tested with data set consisting of 200,000+ records and got the prediction map in under 10s. Subsequently, stress tested with 52 million records data set without crashing the system.

5.2 Overlay Analysis

For this mode of analysis, many ‘semitransparent’ georeferenced digital maps are overlaid with individual transparency control. The problem is not only limited to viewing the combined effect of a set of factors but, rather that of utilizing this information for intelligent analysis. The objective is to address agriculture-specific issues such as the joint effect of the parameters along with display of corresponding legend value under the mouse pointer and comprehension of the agricultural ecology. Figure 13c shows the overlay formation by overlaying the mean solar radiation map (as shown in Fig. 13a) and the agricultural land map (as shown in Fig. 13b).

Unlike the image overlay and merging that is also supported by commercial image editing tools, the proposed overlay analysis as per the framework will use the legend files (Fig. 9), so that the legend values under the mouse pointer are displayed for each of the overlaid maps.

6 Conclusions and Future Work

In this paper, we proposed a multi-disciplinary, integrated framework for predicting the long-term productivity of perennial pasture grasses in the Kingdom of Saudi Arabia. The proposed framework is based on predictive data mining and spatial data warehousing techniques in a GIS environment. The proposed framework can be used for identification of those regions where fodder is not grown but has a high probability of growth success. Other applications include identification of the crop/fodder varieties that can be successfully cultivated and finally analysis of the agricultural ecosystem using different parameters.

As per the King Abdullah’s Agricultural Initiative [23], the policy of direct investment in foreign agriculture is being pursued. That is, cultivating in countries better suited for cultivation of food staples along with availability of resources (land, water and labor), availability of infrastructure and good relations with the Kingdom. The proposed framework is aligned with that policy, as it can be used to help identify promising regions outside the Kingdom.

DSS and ES based on the proposed framework would be required by the policy makers to act pro-actively rather than reactively. The traditional approach of identifying feasible rangeland areas is not only time-consuming, but also expensive. By using the proposed framework, decision makers, researchers, academicians, etc. will be able to identify and evaluate the potential of cropping/fodder regions within the Kingdom, along with computed probabilities of success of

the selected species of crop/fodder. This will help enhance the area under cultivation, increase yield, increase healthy livestock population (camels, sheep, goats, cattle), etc., thus increasing economic prosperity of the Kingdom. The results of the proposed framework are likely to demonstrate the utility of data-driven DSSs for the analysis of pasture grass yields, as simple models are incapable of predicting the outcome of complex yield–meteorological relationships. Thus, our work has the potential to help improve food security of the Kingdom.

Acknowledgments This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH)—King Abdulaziz City for Science and Technology—the Kingdom of Saudi Arabia—Award Number 12-AGR2709-3. The authors also acknowledge with thanks Science and Technology Unit, King Abdulaziz University for technical support.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bédard, Y.; Merrett, T.; Han, J.: Fundamentals of spatial data warehousing for geographic knowledge discovery. *Geogr. Data Min. Knowl. Discov.* **2**, 53–73 (2001)
- Cereal yield (kg per hectare). <http://data.worldbank.org/indicator/AG.YLD.CREL.KG/countries>. Accessed 27 June 2015
- Holland, R.; Loveday, D.; Ferguson, K.: How Much Meat To Expect From A Beef Carcass. University of Tennessee Institute of Agriculture, Knoxville (2015)
- The era of cheap food may be over, <http://www.theguardian.com/business/2012/sep/02/era-of-cheap-food-over>. Accessed 27 June 2015
- KSA remains Australia's largest meat export market in MENA (2013). <http://www.saudigazette.com.sa/index.cfm?method=home.regcon&contentid=20130714173444>. Accessed 27 June 2015
- Trading economics. <http://www.tradingeconomics.com/saudi-arabia/arable-land-percent-of-land-area-wb-data.html>. Accessed 27 June 2015
- Oksanen, J.; Minchin, P.R.: Continuum theory revisited: what shape are species responses along ecological gradients?. *Ecol. Model.* **157**, 119–129 (2002)
- Guisan, A.; Theurillat, J.P.: Equilibrium modeling of alpine plant distribution: how far can we go?. *Phytocoenologia* **30**, 353–384 (2000)
- McCune, B.: Non-parametric habitat models with automatic interactions. *J. Veg. Sci.* **17**, 819–830 (2006)
- Hastie, T.; Tibshirani, R.; Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, 2nd edn. Springer, New York. <http://www-stat.stanford.edu/tibs/ElemStatLearn/main.html>. Accessed 27 June 2015
- Dhanya, C.T.; Nagesh Kumar, D.: Data mining for evolution of association rules for droughts and floods in India using climate inputs. *Geophys. Res. Atmos.* **11**(D2), 1–15 (2009)
- Nadiammal, G.V.; Krishnaveni, S.; Hemalatha, M.: Evaluating the yield of hybrid napier grass with data mining techniques. *Int. J. Comput. Appl.* **35**(5), 1–7 (2011)
- Zhang, B.; Valentine, I.; Kemp, P.: Spatially explicit modelling of the impact of climate changes on pasture production in the North Island, New Zealand. *Clim. Change* **84**, 203–216 (2007)
- Abdullah, A.: Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (online analytical processing) tool. *J. Comput. Electron. Agric.* **69**(2009), 59–72 (2009)
- Uno, Y.: Application of Machine Learning Methods and Airborne Hyperspectral Remote Sensing for Crop Yield Estimation. McGill University, Montreal (2003)
- Siqueira, M.; Paço, T.; Silvestre, J.C.; Santos, F.L.; Falcão, A.O.; Pereira, L.S.: Generating fuzzy rules by learning from olive tree transpiration measurement—an algorithm to automatize Granier sap flow data analysis. *Comput. Electron. Agric.* **101**(0), 1–10 (2014)
- Demšar, D.; Džeroski, S.; Henning Krogh, P.; Larsen, T.: Using machine learning to predict the impact of agricultural factors on communities of soil microarthropods. *AMS* **2**, 147–159 (2005)
- Kamalak, P.; Hemalatha, K.: Agro genius: an emergent expert system for querying agricultural clarification using data mining technique. *Int. J. Eng. Sci.* **1**(11), 34–39. ISSN:2278-4721 (2012)
- Spatial datawarehouse. <http://www.spatial-eye.com/Engels/Applications/Spatial-DWH/page.aspx/117>. Accessed 27 June 2015
- Abdullah, A.: An analysis of Bt-Cotton cultivation in Punjab, Pakistan using the agriculture decision support system (ADSS). *Agbioforum J.* **13**(3), 274–287 (2010)
- Zhang, B.; Valentine, I.; Kemp, P.D.: explicit modelling of the impact of climate changes on pasture production in the North Island, New Zealand. *Clim. Change* **84**, 203–216 (2007)
- Wang, C.; Sang, L.; Yang, J.; Zhang, C.; Zhu, D.; Ming, D.: Spatial identification of connected arable lands using geometric network model. *Sens. Lett.* **10**(1–2), 1–2 (2012)
- Al-Obaid, A.A.; Deputy Minister For Agricultural Research and Development Affairs: King Abdullah s' Initiative for Saudi Agricultural Investment Abroad: a way of enhancing Saudi food security. Presentation at Expert Group Meeting on “achieving food security in member countries in post-crisis world”. Islamic Development Bank, Jeddah (2010)
- Xie, J.; Xu, L.; Chen, E.: Image denoising and inpainting with deep neural networks. In: Pereira, F.; Burges, C.J.C.; Bottou, L.; Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, pp. 341–349. Neural Information Processing Systems Foundation, Inc., Lake Tahoe (2012)
- Barcelos, C.A.Z.; Batista, M.A.: Image restoration using digital inpainting and noise removal. *Image Vis. Comput.* **25**(1), 61–69 (2007)
- Abdullah, A.; Alshamat, S.; Bakshwain, A.; Aslam, A.: Automatic Text removal and replacement in scanned meteorological thematic maps. *J. Comput. Sci.* ISSN:1549-3636 (2015)
- Langley, P.; Iba, W.; Thompson, K.: An analysis of Bayesian classifiers. In: *Proceedings of the 10th National Conference on Artificial Intelligence*, pp. 223–228. MIT Press (1992)
- Papoulis, A.: *Probability, Random Variables, and Stochastic Processes*. Front Cover. pp. 576 McGraw-Hill, New York (1984)
- Alsherif, E.A.; Ayesha, A.M.; Allogmani, A.S.; Rawi, S.M.: Exploration of wild plants wealth with economic importance tolerant to difficult conditions in Khulais Governorate, Saudi Arabia. *Sci. Res. Essays* **7**(45), 3903–3913 (2012)
- Warnock, D.L.; Leep, R.H.; Bughrara, S.; Min, D.H.: Cold tolerance evaluation of improved diploid and tetraploid cultivars of perennial ryegrass. *Crop Manag.* doi:10.1094/CM-2005-0221-01-RS. <http://www.plantmanagementnetwork.org/pub/cm/research/2005/cold>. Accessed 27 June 2015

31. Lamy, E. et al.: Factors Influencing Livestock Productivity. pp. 19–30. Springer, Berlin (2012)
32. Killmann, W.: Climate Change and Food Security: A Framework Document, pp. 1–110. FAO, Quebec City (2008)
33. Alexandratos, N.; Bruinsma, J.: World agriculture towards 2030/2050: the 2012 revision (No. 12-03). ESA working paper (2012)
34. Bruinsma, J. (Ed.): World Agriculture: Towards 2015/2030: An FAO Perspective. Earthscan, London (2003)
35. Linehan, V.; Thorpe, S.; Gunning-Trant, C.; Heyhoe, E.; Harle, K.; Hormis, M.; Harris-Adams, K.: Global food production and prices to 2050. In: Paper Presented at the 43rd ABARES Outlook Conference (2013)
36. Fischer, G.: World food and agriculture to 2030/50. Technical paper from the Expert Meeting on How to Feed the World in 2050 (2009)
37. Larkin, B.: Red meat market report. http://www.mla.com.au/files/17d47c22-ebc4-4520-a890-a1e800aa6f40/RMMR_MENA_June_2013.pdf. Accessed 27 June 2015
38. Beaulieu, D.: Annual plants (2014). http://landscaping.about.com/od/landscapingdictionary/g/annual_plants.htm. Accessed 27 June 2015
39. Scanned Image to Excel Converter. <http://www.verypdf.com/app/scan-to-excel-ocr/scanned-image-to-excel-converter.html>. Accessed 27 June 2015
40. Siqueira, T.L.L.; Ciferri, R.R.; Times, V.C.; de Aguiar Ciferri, C.D.: Benchmarking spatial data warehouses. In: Pedersen, T.B.; Mohania, M.K.; Tjoa, A.M. (eds.) Data Warehousing and Knowledge Discovery, pp. 40–51. Springer, Berlin (2010)
41. Nayab, N.: Disadvantages to using decision trees. <http://www.brighthubpm.com/project-planning/106005-disadvantages-to-using-decision-trees/>. Accessed 27 June 2015
42. Auria, L.; Moro, R.A.: Support Vector Machines (SVM) as a Technique for Solvency Analysis. Deutsches Institut für Wirtschaftsforschung (DIW), Berlin (2008)
43. Linear regression model, CAMO. http://www.camo.com/rt/Resources/linear_regression_model.html. Accessed 27 June 2015
44. Ross Quinlan, J.: C4.5: Programs for Machine Learning. ISBN:1558602402 (2014)
45. Zhao, Y.; Zhang, Y.; , : Comparison of decision tree methods for finding active objects. J. Adv. Space Res. **41**(12), 1–14 (2007)
46. Pearl, J.: Bayesian decision methods. In: Shafer, G.; Pearl, J. (eds.) Readings in Uncertainty Reasoning, pp. 345–352. Morgan Kaufmann, San Mateo (1990)
47. John. C.: Raster to vector conversion in a local, exact and near optimal manner (1991). <https://www.dwa.gov.za/iwqs/r2v/main.shtml>. Accessed 27 June 2015
48. Anguita, D.; Ghio, A.; Ridella, S.; Sterpi, D.: K-fold cross validation for error rate estimate in support vector machines. In: Proceedings of DMIN, pp. 291–297 (2009)