Agricultural Data Visualization for Prescriptive Crop Planning

Harshitha B P, Amith R, Abhishek S, Rohit Vibhu C

Student & Computer Science and Engineering, SJB Institute of Technology
Bangalore, Karnataka. India.

Abstract— *The importance of carrying out effective and sustainable agriculture is getting more and more obvious. Modern computers and the internet have made it much easier to make graphics out of tabular data and give these graphics the qualities of animation and interactivity through data visualization.*

Investigate effects of climate change on crop yields over time, climate change effects on crop yield eclipsed by other developments (technology, crop mix change), help make informed decisions for sustainable growth. Agriculture yield data is used to analyze and improve the crop yield and represent in the form of a Graphs through data visualization technique. The visualization methods presented include interactive charts to enable our data users to drill down and focus on more detailed views of these data displays. Each of these methods facilitates the display of large volumes of data and allows data users to extract information from our statistics that is difficult or impossible to obtain from traditional static charts or tabular displays of data.

Keywords - *Data Visualisation, Predictive Crop yield.*

I. Introduction

Well-designed data graphics and maps can help someone understand the underlying data much better than a simple table of numbers. Until recently, historical data offerings on the NASS website were generally limited to static charts and maps, and text and tables via text files, spreadsheets and PDF documents**.** Data attributes such as crop yield measurements, area of farm land, share of crop types, temperature, precipitation, fertilizer, etc. Rich dataset allowed us to consider many options. Given the rich and comprehensive nature of NASS historical data, it is important to effectively and efficiently present this information. Charts can quickly become cluttered with too much data, but good graphical design can reveal new insights from these data.

There has been a recent surge of interest in data visualizations and their potential to communicate effectively. This rise has been influenced by the increasing availability of tools for creating visualization, and the spike in data visualization use. News outlets around the globe have been at the forefront of this charge, experimenting with unique and appealing ways to present information to the public. Their investments and dramatic outputs have provoked questions amongst other industries about what benefits data visualizations could contribute to communicating information more effectively.

Precision agriculture (PA) and information technology (IT) are closely interwoven. The former usually refers to the application of nowadays' technology to agriculture. Due to the use of sensors and GPS technology, in today's agriculture many data are collected. Making use of those data via IT often leads to dramatic improvements in efficiency. For this purpose, the challenge is to change these raw data into useful information. Techniques or methods are required which use those data to their full extent – clearly being a data mining task. Self organizing maps and multidimensional scaling techniques will be used to reduce the high-dimensional input data to two dimensions. The processed data can then be visualized appropriately on 2D maps. An analysis of correlations and interdependencies in the data set will be given, based on the visualization.

The agricultural statistics is the main backbone for any visualization process to start with, this agricultural statistics series will be based on the Global Strategy to Improve Agricultural and Rural Statistics that is located in the resource section and available from the web page . The Global Strategy is based on input from national, regional, and international institutions that have an interest in improving agricultural statistics to meet growing user needs for reliable, timely, relevant, comparable, consistent, and accessible agricultural data. There is an increasing focus on the role of agricultural in rural development, where many of the world's poor live and work. These data are also needed for monitoring of the Millennium Development Goals on a country, regional and international basis. The Global Strategy cites a decline in the quality and suitability of agricultural data for these purposes. This module will provide an overview of the importance and uses of agricultural data to understand the current and emerging needs of data users and broaden the vision for agricultural statistical systems.

Statistically speaking, agricultural statistics are seen as part of the economic profile of a country but are increasingly becoming more broadly defined as data on agriculture in relation to the environment, climate change, biodiversity, food security, and natural resource management are being examined. In many countries the relationships between the ministries of agriculture and national statistical organizations could be improved in terms of integrating agricultural data with other available data to provide a more comprehensive picture of the issues within a country.

We can look at agricultural data in two general groups:

1. The structure of agriculture in a country, which includes agricultural holding by distribution, size, tenure, land use, means of production and labor force.

2. Annual agricultural activities which includes crop and livestock production, trade and prices of agricultural products and labor force information.

Increasingly Agricultural data is seen in the broader context of social, environmental and economic needs of the population. As populations increase so does their demand for agricultural products and need for economic stability. Therefore agriculture is moving from simply focusing on production to sustainably meeting social needs without damaging the environment while remaining economically viable to the mostly lower income population. Globalization means that raw materials can be obtained from one country, processed in a different country, and marketed in virtually all countries. Advances in technology, transportation and communication have made it more effective to conduct business where it is geographically advantageous rather than being limited by the rules, regulations and resources where the business is located.

International organizations and increasingly national governments need regional and global data for planning and estimation of future needs and impacts. Agricultural activities and policies in one country affects other countries through global trade and environmental impact. Agricultural statistics are needed to provide information used to monitor trends and estimate future prospects for agricultural commodity markets which can assist in setting policies such as price supports, where the price of a commodity is artificially maintained by government action to stabilize the economy, or tariffs, a fee charged by governments on particular goods that are either imported or exported. Agricultural statistics are also needed to assess the role of agriculture commodity production in trade and economic development, not only at the national level but also at the regional level.

The world population is increasing as the amount of land suitable for cultivation is diminishing, leading to food security problems. Data is needed for international organizations to analyze the food security situation and plan efforts to meet agricultural production requirements. There is an increasing interest in examining the environmental impact of agricultural activities including issues such as chemical use, genetic engineering, biodiversity, water conservation and land use.

National governments need information for development planning, especially in rural areas where a large percentage of the population may be economically dependent on agriculture. Agricultural statistics provide this information by holding, type and area, which is not available from other data sources in the detail required. These data on agricultural commodity production form part of the System of National Accounts (SNA).

Combining the data on availability and access with weather and crop predictions, as well as tracking historical natural disaster patterns, can help estimate areas of future food insecurity for planning purposes. Data can be used to analyze a particular crop or livestock of interest, such as the main export for the country for example, to track changes and look for ways to maximize production. Studies could be carried out by geographical area to look at land use and agricultural production and see if there are better ways to maximize production within the particular ecological zone while preserving natural resources. For policy purposes it would be useful to analyze data by various types of agricultural holdings such as subsistence versus market-oriented and we've already discussed by gender.

Agricultural censuses are necessary to improve intercensal survey estimates and provide detailed small area data that is not available from survey data. In order to do this, data must be comparable, which means standardized definitions, unit of enumeration, and data collection methodologies. Complete documentation can also assist in determining the potential causes of errors in estimates and help in determining ways to improve the estimates.

II. PROBLEM STATEMENT

The idea is to develop an application that analyses data sets of temperature, rainfall and reservoir levels for a region in Karnataka, and project a visualization that statistically determines the viability of crops that can be grown in the selected region.

III. LITERATURE SURVEY

There has been a recent surge of interest in data visualizations and their potential to communicate effectively. This rise has been influenced by the increasing availability of tools for creating visualization such as maps, and the spike in data visualization use. News outlets around the globe have been at the forefront of this charge, experimenting with unique and appealing ways to present information to the public. Their investments and dramatic outputs have provoked questions amongst other industries about what benefits data visualizations could contribute to communicating information more effectively. While in the past data visualization was viewed as an important analytical tool for researchers, it is quickly being recognized as an essential aspect of effective research communication. Although data visualization is fairly new for development researchers, it affords opportunities to both transform and display data(Crop yield data)[3] . Visualization proponents also highlight that these capabilities are extremely useful within complex and changing environments, which are akin to the contexts surrounding IDRC-supported projects. Data visualization gives precise statistical information for self organizing maps[2].

The importance of data visualization is further heightened by the increasing digitization of the world, which has created information-overloads in a time-deprived policy and development sector. For an example of various managing strategies, see e.g. [5].Our approach of using SOMs is motivated by the need to better understand the available yield data and extract knowledge from those data. SOMs have been shown to be a practical tool for data visualization [4]. One of the strengths of IDRC-supported projects has been the under-researched regions and fields that are explored. However, while the collection of this information is often groundbreaking and innovative, the expounded findings still have to be heard within saturated information markets. The utility of this research is therefore dependent on how it is communicated and the level of interest and investment from stakeholders and policymakers. This study assesses the potential of data visualization to assist in effectively communicating research for influence.

With the development of the global positioning systems (GPS), geographical information software (GIS) and various sensors and actuators, the possibility of initiation information-guided plant production is never been greater. The assumption is that the more information and precision is put into cultivation management, the higher the profit will be. During recent years, there has been no significant increase of precision agriculture (PA), with only a low share of farmers employing PA applications. However, there are many suppliers of PA hardware and software[5].

Data visualization or data visualisation is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information".[1]

A primary goal of data visualization is to communicate information clearly and efficiently via statistical graphics, plots and information graphics. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message.[2] Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships in the data for one or more variables.

Data visualization is both an art and a science.[3] It is viewed as a branch of descriptive statistics by some, but also as a grounded theory development tool by others. Increased amounts of data created by Internet activity and an expanding number of sensors in the environment are referred to as "big data" or Internet of things. Processing, analyzing and communicating this data present ethical and analytical challenges for data visualization. The field of data science and practitioners called data scientists help address this challenge.[4]

Today, the world is experiencing another surge in data visualization popularity. This interest can be partially linked to the increased availability of new technologies and software products which enable every user to dabble in the world of visualization. However, these resources have not come about on their own, but have been the by-products of years of research and development from an international community of scholars and practitioners. In a recent publication Evert Lindquist examines visualization by parsing the field into three unique disciplinary streams: information visualization, graphics and information display, and visual facilitation for thinking and strategy (Lindquist, 2). While each of these streams is distinctive in both their approach and focus; there are larger overlaps which undercut any hard fast boundaries. That said, exploring these three streams provides a stronger understanding of the rich and diverse scholarship which has contributed to the field of data visualization.

A collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization

Graphics and information display is the first stream of the visualization, which focuses on the aesthetics of displaying information graphically, rather than enabling the data to determine the form. Lindquist's summary of this area highlights that there is an astounding diversity of approaches; covering everything from designing algorithms to enable visualization production, to understanding cognitive interpretations of different graphical forms, to exploring the applications and theoretical constructs of data visualization (Lindquist, 2011). Overall, what unites the divergent approaches of this stream is a concentration on the design of visualizations and how form can strengthen utility for purposes of communication, marketing and illumination[6].

Human perception/cognition and data visualization

Almost all data visualizations are created for human consumption. Knowledge of human perception and cognition is necessary when designing intuitive visualizations. [14]Cognition refers to processes in human beings like perception, attention, learning, memory, thought, concept formation, reading, and problem solving.[15] Human visual processing is efficient in detecting changes and making comparisons between quantities, sizes, shapes and variations in lightness. When properties of symbolic data are mapped to visual few properties, humans can browse through large amounts of data efficiently. It is estimated that 2/3 of the brain's neurons can be involved in visual processing.[16]Proper visualization provides a different approach to show potential connections, relationships, etc. which are not as obvious in non-visualized quantitative data. Visualization can become a means of data exploration.

The concept of precision agriculture first emerged in the United States in the early 1980s. In 1985, researchers at the University of Minnesota varied lime inputs in crop fields. It was also at this time that the practice of grid sampling appeared (applying a fixed grid of one sample per hectare). Towards the end of the 1980s, this technique was used to derive the first input recommendation maps for fertilizers and pH corrections. The use of yield sensors developed from new technologies, combined with the advent of GPS receivers, has been gaining ground ever since. Today, such systems cover several million hectares.

In the American Midwest (US), it is associated not with sustainable agriculture but with mainstream farmers who are trying to maximize profits by spending money only in areas that require fertilizer. This practice allows the farmer to vary the rate of fertilizer across the field according to the need identified by GPS guided Grid or Zone Sampling. Fertilizer that would have been spread in areas that don't need it can be placed in areas that do, thereby optimizing its use.

Around the world, precision agriculture developed at a varying pace. Precursor nations were the United States, Canada and Australia. In Europe, the United Kingdom was the first to go down this path, followed closely by France, where it first appeared in 1997-1998. In Latin America the leading country is Argentina, where it was introduced in the middle 1990s with the support of the National Agricultural Technology Institute. Brazil established a state-owned enterprise, Embrapa, to research and develop sustainable agriculture. The development of GPS and variable-rate spreading techniques helped to anchor precision farming[16] management practices. Today, less than 10% of France's farmers are equipped with variable-rate systems. Uptake of GPS is more widespread, but this hasn't stopped them using precision agriculture services, which supplies field-level recommendation maps

The impact of digital agriculture on the field is well documented and researched. From variable rate application to realtime NDVI visualization (i.e., the index for visualizing vegetation health), farming will forever be changed. In the future, data creation, analysis and decision-making will almost certainly increase at the field level. Farming operations will have the opportunity to prosper from targeted field solutions, data-driven agronomic advice and smarter inputs. Software is being developed to help propel developing countries toward modern farming practices.

While the first two revolutions in agriculture — mechanization and biotech — had a major impact for farmers and select agribusinesses, digital agriculture will fundamentally transform every part of the agribusiness value chain. Seed companies were not drastically changed from the bottom up to accommodate advanced machinery. While significant innovation occurred in equipment design, it was not altered to specifically accommodate genetically modified seeds. However Ag 3.0 will affect producer buying behavior and seed and equipment product design, and could enable dynamic pricing at the consumer retail level.

These implications will gradually affect multiple business functions across a single company. For example, digital agriculture and big data will change the way seed and agrichemical companies market, price and sell products, select and invest in their R&D pipeline, recommend and technically support product sales, manufacture and distribute products, and manage credit and financial risk. Business strategy, product design, customer preferences and even organizational structure will change as more digital agriculture data is available. As technological advancements in equipment and inputs slow, companies will need increasingly to compete on digital strategy. Being able to support digital agriculture becomes ever more important.

This revolution will also challenge traditional company roles, intercompany relationships, reward systems and, potentially, entire business models. Digital agriculture is creating competition among both traditional and nontraditional competitors. The industry is in a storming phase and agribusinesses are working to solidify their place in Ag 3.0. Several companies are investing heavily in internal data activities such as standardization, storage, software and analytics. Others are focusing on outsourcing strategies or licensing software from other companies. Still others are taking a wait-and-see approach. As the industry evolves, disruption will follow. It is essential for agribusinesses to transform their business and themselves to differentiate and provide more value to customers.

Farms are consolidating at an increasing rate as technology supports automation and economies of scale. Input applications are based on factual data and investments into farming tech are funded by profit saved by data-driven efficiency. While the benefits of digital agriculture are compelling, it has been met by significant challenges, for example, difficulty using software, data usage concerns, disparate and propriety data formats and an unclear return on investment. Agribusiness has struggled to provide immediate, tangible results from digital agriculture equipment and software.

IV. DESIGN & IMPLEMENTATION

The visualization application takes into account the following data sets for the present and the next year:

- Temperature datasets for the present and next year
- Rainfall data sets for the present and next year
- Reservoir water levels and discharge quantity

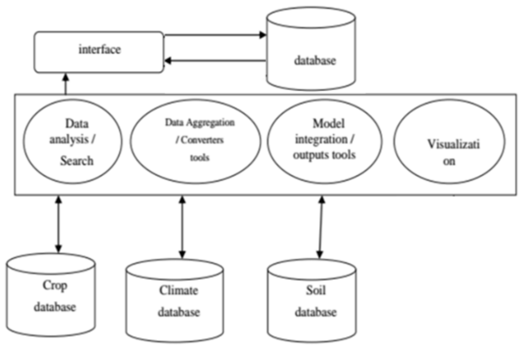
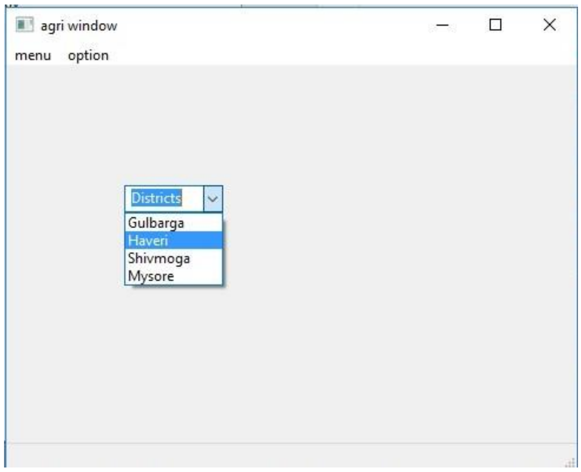


FIG. 1.2 Block Diagram of the Data visualization toolkit

These datasets can be comma separated values or excel spreadsheets which contain the temperature readings and rainfall in cms and reservoir discharge quantity information in each month for the present and following year .The information would be read directly from these excel sheets. The duration would be chosen based on the month of information publication and the statistics would be visualized based on the average duration the crop takes for its complete growth and the temperature, water availability statistics.

Interface

The interface provides a facility to choose the region, an option to browse and choose the data file and the option to choose the month from which the cultivatable statistics have to be taken into account and visualized. This is being designed using PyQt and QtDesigner which is an interface development



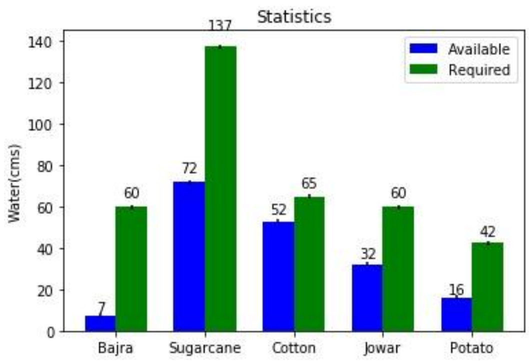
module.

Fig 1:

The above figure shows the selection being made on the interface window. The interface consists of 2 windows which give general information of agricultural statistics and also a selection window that consists of a drop down menu. The region is selected from the list and the visualized statistics are obtained on another pop-up window.

Visualization Process

The Visualization process consists of three levels of filtering the viability of crops. The first phase would be elimination of non-cultivable crops based on soil types present in the region selected. The second phase involves filtering of crops according to temperature statistics. The third phase involves analyzing water availability (taking into consideration both rainfall statistics and reservoir outflow levels) and filtering the crops accordingly. The entire data thus obtained would undergo the final phase which calculates the extent of drift of the available information from the prescribed or required values. More the drift less is the probability



of growth and survival of crops and vice versa.

Fig 2.

The above figure shows the analyzed results of water statistics as a bar graph visualization showing the available quantity of water (in cms) and the required quantity for cultivation of each crop. If the available value exceeds the required value then it indicates that the crop can grow viably.

Results

The data thus analyzed would be projected and visualized using bar graphs and pie charts that indicate the required and available values of water availability and temperature readings. The ratios of the available and required values would indicate the probability of growth. If the value is nearer to 1, that particular crop would have the highest chances of cultivation.

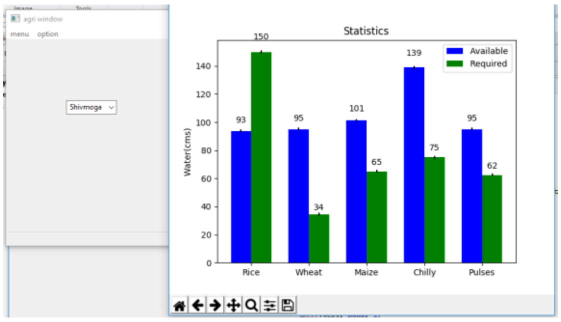


Fig 4.

The above figure shows the visualization based on the temperature readings along with the probability ratios for each crop. The input for the visualization is the region name selected from the drop down menu in the interface of the application which here, is selected as “Shivmoga”. The value nearest to the value "1" represents the crop with the higher chances of sustainable cultivation. The difference between the requirement and availability is seen as a measure of success rate of the crop. Higher the difference, then more is the probability of growth and vice versa.

V. Conclusions

This survey aims at promoting smart farming mechanisms by prescribing possible crops by incorporating statistical data visualization techniques. This would help famers plan their cultivation in a viable manner based on the regional parameters. This promotes prescriptive crop cultivation based on smart farming and digital analysis methods. The key role of this project is to promote more success rates of cultivation and prevent losses arising from unplanned cultivation while also benefiting the farmer with commercial profits. Thus it helps in establishing a balance of growth and profit rates of crop cultivation.

The charts and maps developed and presented in this paper were produced using a variety of products and programming languages. These programs and products are all useful technologies, but they are most useful when information designers think critically about how best to display the data of interest. When looking at historical data, researching and applying best principles and appropriate technologies to create well-designed graphics can help data-users find the information they want, and learn more about the big picture, and the story behind the data.

References

[1] ([1] NASS Research, Science and Technology website. (http://www.nass.usda.gov/Research_and_Science/)

[2] [2] Timo Honkela, Samuel Kaski, Krista Lagus, and Teuvo Kohonen. WEBSOM—self- organizing maps of document collections. In Proceedings of WSOM’97, Workshop on Self- Organizing Maps, Espoo, Finland, June 4-6, pages 310–315. Helsinki University of Technol- ogy, Neural Networks Research Centre, Espoo, Finland, 1997. 2.

[3] [3] T. Jensen, A. Apan, F. Young, and L. Zeller. Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform. Comput. Electron. Agric., 59(1-2):66– 77, 2007.

[4] [4] Timo Honkela, Samuel Kaski, Krista Lagus, and Teuvo Kohonen. WEBSOM—selforganizing maps of document collections. In Proceedings of WSOM’97, Workshop on SelfOrganizing Maps, Espoo, Finland, June 4-6, pages 310–315. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, 1997.

[5] [5] M. Schneider and P. Wagner. Prerequisites for the adoption of new technologies - the example of precision agriculture. In Agricultural Engineering for a Better World, Dusseldorf, 2006. ¨ VDI Verlag GmbH.

[6] [6] Lindquist, Evert. “Grappling with Complex Policy Challenges: Exploring the Potential of

[7] Visualization for Analysis, Advertising and Engagement”, *Discussion Paper for HC Coombs Policy Forum Round tables, (2011).*

[8] [7] Jacqueline Strecker, Research Award Recipient IDRC; & Tricia Wind, Evaluation Officer IDRC. Report author: Jacqueline Strecker Feature review by: Amanda Cox Winter 2012

[9] [8] Tufte, Edward R., *Envisioning Information* (Chesire, Conn.: Graphics Press, 1990).

[10] [9] Iliinsky, Noah, “On Beauty”, Ch. 1 in Steele and Iliinsky (eds.), *Beautiful Visualization*

[11] (Cambridge: O’Reilly, 2010), pp.1-13.