BIG DATA
CASE Study

WORLD BANK GROUP

Data-driven agronomy will help increase the sector's capacity to adapt to climate change

Mining Big Data for Climate-Smart Agriculture

Erick C.M. Fernandes, Daniel Jiménez, Andy Jarvis and **Sylvain J. Delerce**

SUMMARY

Climate change means traditional calendar-based decisions about what and when to plant are no longer reliable for farmers. New approaches are urgently required to provide relevant and timely information for decision making. In Colombia, the International Center for Tropical Research (CIAT) has combined long-term data on rice harvests and climate patterns, and analyzed it to provide essential information for crop planting decisions. This project draws on the Colombian experience. The team partnered with rice-growing experts from Argentina, Brazil, Chile and Uruguay, helped them prepare cropping and climate data, and held a workshop to analyze their data. The technique combines the two databases on harvests and weather, and relates each harvest to the corresponding climate sequence for approximately 120 days between sowing and reaping. The data is analyzed to unravel underlying correlations between climate factors and yield variability.

This enables identification of climate sequences that are favorable or unfavorable for cropping in specific areas. By matching these with seasonal forecasts, the system can advise farmers on what variety of rice to plant for the anticipated conditions, and when to plant it. To support field data capture by farmers, the project also created a web platform and an Android app. The tool can be applied to any crop or location, providing farmers with data-driven information to help them decide what, when and where to plant.

CHALLENGE

Greater weather variability due to climate change means calendar references are increasingly outdated in helping farmers make the right planting decisions. New approaches are urgently required to provide them with relevant information and enhance their resilience to climate variability. Recent climate change analysis suggests Latin America and the Caribbean could have a future climate more suitable for growing rice – the world's most important food crop in terms of energy consumed by humanity. In Colombia rainfall and temperature extremes have changed differently in each region, and the climate is increasingly unpredictable, resulting in national average rice yields falling from six to five tonnes per hectare in less than five years (without noticeable changes in soil or crop management). Faced with an increasingly volatile climate, farmers need new methods for deciding when and what to plant.

This would foster data-driven agronomy and increase the sector's capacity to adapt toclimate change

INNOVATION

CIAT's technique combines and analyzes two country-wide databases, covering commercial harvests and weather. Data on harvests has been collected by the Colombian National Rice-Growers' Federation for almost 20 years, covering variables such as yield, grain humidity, sowing and harvest dates, cultivar, municipality and cropping system (irrigated or upland rice). Data from the Colombian National Meteorology Institute provided daily records of five variables: Maximum and minimum temperature, precipitation, relative humidity and solar radiation.

By combining these two databases, each individual harvest event can be related to a corresponding climate sequence for around 120 days between sowing and harvest. Data is analyzed using machine learning techniques such as Artificial Neural Networks (ANN), random forest and clustering, to reveal the combination of climatic factors that result in high or low yields in a specific region.

Based on the harvest events database for two municipalities, analysis of the 120 days from sowing to harvest for the five climate variables revealed 17 different climatic sequences. By matching these against the yields obtained under each sequence and evaluating the relevance of specific climate indicators against the growing phases of the crop, CIAT identified favorable and non-favorable climatic situations. Projecting these onto seasonal forecasts, the system can advise farmers on optimum sowing dates and rice varieties, given the weather ahead. By providing farmers with data-driven information for planting decisions, the tool can compensate for the increasing irrelevance of traditional knowledge due to climate variability.

Expanding the project reach

In pilot tests, analysis of crop data together with weather data generated advice against planting rice that season, due to projected adverse climate conditions. Farmers received the planting advice via the growers' association. Those who ignored it harvested nothing and lost considerable inputs, while those who followed it saved seed, labor, fertilizer and water. The historical data allowed insight at a site-specific level. For example, in Saldaña – where rice is grown year-round, due to irrigation policies – the analysis showed that rice yields were limited mainly by solar radiation during the grain-ripening stage. This implies that farmers can boost yields by aligning sowing dates with sunnier seasons, or choosing crop varieties resilient to low solar radiation.

With this success suggesting other countries and crops could benefit from the same approach, the project worked to scale it up in Latin America. The team visited Argentina, Uruguay and Brazil to seek partnerships with agricultural associations and assess data availability. CIAT then helped partners with data collection and preparation for five months before holding a workshop in Uruguay to teach them to prepare and analyze their own data. The training involved participants using the system for themselves, to ensure they could replicate the process in their own contexts.

Building a real-time system

Equally important was the move from using only historic data on weather and yields, to collecting real-time data and further automating the analysis to build a real-time system. To enable field data capture by farmers for factors such as soil type and crop management techniques, the team developed an Android app, using open-source software and an external provider to build the app. Supporting a wide range of Android versions, it can capture GPS coordinates using one button. Stored in a cloud, the data is instantly available and can be exported to any format for reuse. The app is linked to a web platform which generates personalized reports, real-time information, interactive graphs and mapping, allowing for optimum site-specific crop management. The system will be further refined by increased use of machines such as drones or inexpensive wireless sensors for data capture. This allows high frequency measuring, without requiring effort from farmers and eliminating the margin of human error.

RESULTS

The workshop teams spent most of the time preparing their data: standardizing and cleaning it, and relating crop data with weather series. This reflects the reality of work with data. Each team completed at least one analysis using random forest models that allowed them to identify the most relevant factors behind yield variability, and evaluate the relationships between input and output variables.

The group quickly grasped the approach. Some teams achieved satisfying results which coincided with previous work, while others obtained unexpected results due to incomplete and insufficiently processed datasets.

All experienced the problems that occur while handling real data (such as outliers or variables out of the range when cleaning data).This generated a rich debate around how best to capture data. The limited capacity of participants' personal computers restricted the exercise to small datasets using only ANN and random forest techniques, prompting CIAT to run the analysis on its more powerful server facilities, with whole datasets. Exploratory results using information from one rice mill showed that previous land use also influences rice production.

The app is being finalized after initial user feedback reported that it was too slow loading and saving information, and did not work on all platforms (IOS/Apple, Windows, etc.). This caused some frustration among users (who wanted quick responses to problems, as if from a service provider rather than a research project).

The system could potentially be adapted to almost any crop, and the team aims to pilot it in Africa in 2016. It is hoped the workshop will form the basis for the establishment of a community of practitioners in big data in agriculture. This would foster a culture of data- driven agronomy and increase the sector's capacity to adapt to climate change.

LESSONS LEARNED

The scaling-up and move towards real-time data capture revealed that data quality is of paramount importance to the reliability of the system, but human relationships through local partnerships and a user community will also be crucial to its success.

- ***Work specifically to create a positive user experience***
  User experience is central to success. Technical issues undermine user trust, making the adoption process harder. A tool must offer sufficient services to engage users, and be easy to operate, otherwise it will not be used.
- ***Invest in regional and local partnerships Partnerships***
  on the ground are key to scaling up – particularly with agricultural organizations, which can source data from numerous locations and deliver recommendations to many farmers efficiently. It's important to gain credibility with partners, so they will share information: Open data-sharing is still in its infancy in many places and primary data holders often have legitimate concerns about how information they share will be used.
- ***Capture, prepare and store data meticulously***
  High-quality input data is essential for accurate results. Agricultural data is readily available, but there is big potential to modernize the methods used to capture, store and share it – for example, through machine-based data sourcing. The data preparation step (addressing missing data, outliers, correlated variables, etc.) must be completed, and organizations should evolve towards cloud-based technologies, so data is centralized and always available.
- ***Create a community of practitioners***
  Networking will be powerful in promoting uptake and refinement of these methods. CIAT hopes to support a user community for data mining techniques in agriculture, similar to the 'R community' – a global network of more than 2 million users and developers who voluntarily contribute technical expertise to maintain and extend the 'R software' language. This approach demonstrates how a tool can be continuously improved.

www.open-aeps.org:8080/