



Monsanto breeding lab: St. Louis, Mo., facility.

Photos courtesy of Monsanto.

Data science at Monsanto

Data-driven analytics in agriculture delivers benefits for farmers, consumers and the planet.

By Shrikant Jarugumilli

The Food and Agriculture Organization (FAO) of the United Nations estimates that the global population will reach 9.1 billion by 2050 and that world food production will need to increase by 70 percent to ensure adequate food security. The FAO also estimates that in order to meet today's needs, agriculture utilizes 70 percent of the planet's available fresh water. These two realities mean that in the future we will have to increase food production while decreasing resource usage – including land and water. At Monsanto we are leveraging plant biotechnology, plant breeding and chemistry in conjunction with the power of data science to develop better agricultural products that optimize field production for farmers, increase the sustainability of global agriculture, and run our internal business operations more efficiently. Founded in 1901 in St. Louis, Mo., Monsanto operates in 60 countries across six continents with more than 350 facilities, over 20,000 full-time employees and net sales of \$14.6 billion in 2017 providing agricultural products. Monsanto's mission is to develop integrated solutions enabling farmers to grow crops while using energy, water and land more efficiently.

All About the Roundtable

From R&D to Harvest:

The Development Journey of a Strong Crop

Monsanto has invested heavily in technologies that are enabling pervasive application of data science across the organization. These investments include high-throughput sequencing technologies, sensors, mobile and cloud computing, and technical platforms for collaborative data science initiatives. In many ways, Monsanto has transformed itself from a “seeds and seed traits” company to a digital agriculture company and is operating with velocity, teaming, focus, reproducibility and manageability as the key drivers to achieve this transformation. Monsanto’s products include agricultural seeds in high-performing germplasm, biotechnology-derived traits that enhance plant productivity, crop protection and digital agriculture tools. Backed by the power of data science, these products are offered as integrated solutions to address the full spectrum of challenges a farmer must overcome to deliver a successful crop.

From R&D to Harvest: Journey of a Strong Crop

Our seed innovations begin in R&D, then are rigorously tested to ensure compliance with regulatory and industry standards, are manufactured, and then distributed by our supply chain before being sold to farmers. The foundation of successful seed products are the right genetics bred into elite germplasm for crop qualities such as good yield, strong stalks and roots, and disease tolerance. Often, these products are genetically modified with multiple biotechnology-derived traits to give them above- and below-ground pest protection, to make them resistant to one or more herbicides, or more tolerant to drought conditions. Finally, seeds may be coated with treatments that will help the plants succeed underground by giving them mildew or below-ground pest resistance, or to help their roots better utilize nutrients in the soil.

During the growing season, the farmer will be able to control yield-robbing weeds and pests by applying appropriate crop protection products. Throughout the growing season, Monsanto’s on-farm digital agriculture tools, commercialized by our subsidiary The Climate Corporation, guide the farmer to maximize yield using insights gained by the analysis of multiple data sets that help inform the more than 40 decisions for crop management that are made between planting and harvest. These include decisions about product planting (placement and density) and water and fertilizer use (which parts of the field and in what amount).

All of these offerings are based on understanding the needs of customers, which vary from field

The Roundtable consists of the institutional members of INFORMS with member company representatives typically the overall leader of O.R. activity. The Roundtable is composed of about 50 organizations that have demonstrated leadership in the application of O.R. and advanced analytics. The Roundtable culture is peer-to-peer, encouraging networking and sharing lessons learned among members.

The Roundtable meets three times a year. Roundtable goals are to improve member organizations’ OR/MS practice, help Roundtable representatives grow professionally and help the OR/MS profession to thrive. Further information is available at <http://roundtable.informs.org>.

The Roundtable also has an advisory responsibility to INFORMS leadership. According to its bylaws, “The Roundtable shall regularly share with INFORMS leadership and advise the INFORMS Board on its views, its suggested initiatives and its implementation plans on the important problems and opportunities facing operations research and the management sciences as a profession and on the ways in which INFORMS can deal proactively with those problems and opportunities.” The Roundtable Board meets with the INFORMS Board each spring to discuss topics of mutual concern.

This series of articles aims to share with the INFORMS membership at large some information and insights into how O.R. is carried on in practice today.

to field, and are powered by the application of data science across the company.

Machine Learning Drives Plant Breeding Transformation

We have applied machine-learning algorithms to transform our plant breeding pipeline by leveraging 15 years of data from a corn R&D pipeline to move early product testing from the field to the lab, which is helping our researchers more precisely predict how thousands of seeds will perform in the field. This approach enables us to evaluate roughly five times more corn varieties without having to plant them in fields each testing season and accelerates the delivery of new products by a year. In addition to the development of a machine-learning model, this effort required creation of entire systems of analytics to optimize genotyping lab throughput, ensure data quality, impute full genome genetic information from partial observations and guide advancement decisions based on the genotypic data.

Genetic data drives many critical decisions that advance products through Monsanto’s breeding pipelines. A key example is genome-wide selection (GWS), which provides performance predictions based on genetics. Executing a GWS model requires observing (genotyping) a large number of genetic locations (markers) for each seed, which is cost prohibitive at the scale we require. Fortunately, knowledge about a seed’s ancestors and siblings can be used to reduce the number of direct observations needed. For a corn seed, we typically observe around a thousand genotypes directly. Using a method known as genotype imputation, we can commonly produce high-quality, quantified estimates for another 50,000 markers.

Genetic data drives many critical decisions that advance products through Monsanto’s breeding pipelines.

Roundtable profile: Monsanto



Spring planting: Monsanto's Monmouth, Ill., facility.

Photos courtesy of Monsanto.

IoT Enables Global Field Testing Excellence

Monsanto's Internet of Things (IoT) platform collects data from sensors deployed in fields, on various equipment, and on machinery within processing facilities to provide a high volume and variety of data. Peak data collection for the IoT platform can hit hundreds of millions of data points per day providing invaluable information across thousands of field trials globally.

In addition to this high-velocity data, we have petabytes of genomic data for our seeds coupled with hundreds of millions of acres of associated environmental information to form a rich backbone for data modeling to determine which seed genetics will be most successful in each location. Data scientists apply predictive, prescriptive and cognitive analytical techniques to these data sets. These analytical techniques coupled with our elastic cloud computing capabilities are enabling the derivation of never-before-possible insights that are delivering tremendous value to our customers.

Smarter Supply Chain and Customer Experience

Monsanto leverages a variety of optimization and simulation techniques to increase the efficiency of operational activities and drive innovations and inventions throughout the company. These applications can be classified into two categories: planning for our facilities and planning for our field

operations. Our facilities include laboratories, greenhouses, seed manufacturing and seed packaging units. For efficient utilization of these facilities, we have developed several optimization and simulation models for determining medium- and long-range plans. Typical optimization and simulation models support field operations (cover planting, in-season and harvest operations) that are tailored across the business functions of R&D and supply chain. These models need to be adapted for technological, regulatory and other constraints specific to regions.

It is often difficult to predict seed production yield of new products from their years of available research trial data because research trials and seed production are performed

in different fields. The ability to classify environmental zones within fields and relate them across fields allows for accurate seed production yield estimates derived from research trials. The historical analysis of trial data has made it evident that yield performance depends on the interaction between the crop genotype and environmental factors such as topography, soil and climate conditions at the location where the crop is planted, that is, the zones within the field.

By applying advanced geospatial modeling and mathematical optimization in our supply chain, we can better understand, model and predict the impact that genetics, environmental conditions and agronomic practices have on crop yield. Understanding this relationship helps to mitigate product supply risk by reducing the variability between how much we need to produce and how much is produced by fields. Furthermore, such insights enable Monsanto to reduce its seed production footprint, supporting our ongoing commitment to operating as sustainably as possible.

This work has resulted in a significant increase in seed production yield just by planting a product in the best zones within our existing seed production network. The yield increase can be used either to reduce production acres or the level of uncertainty within existing production acres.

The supply chain optimization system drives an integrated, optimized solution through the

Plan, Source, Make and Deliver functions. Past process improvement initiatives focused on improving the siloed processes that are required to deliver supply chain excellence. This effort led to very efficient groups within the overall supply chain system, but often sacrificed overall efficiencies by suboptimizing the overall supply chain.

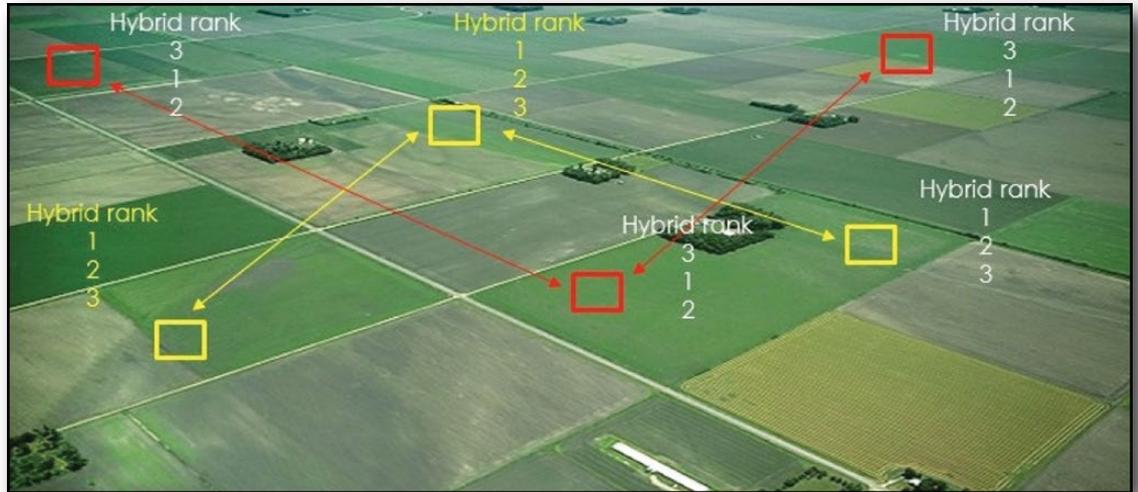
The goal of the new system was to drive an integrated approach to customer satisfaction through supply chain excellence utilizing optimization, thereby allowing for the true integration of the processes, constraints and costs that challenge the business. This initiative required not only new systems to support the operations research implementation, but also a new business mindset. The understanding of the potential cost increases and process changes within one function, which appear suboptimal when viewed in a silo, needed to be understood as they relate to the overall efficiency of the end-to-end supply chain.

As key areas of the business underwent a procedural transformation and education effort to drive efficiencies, the optimization and systems groups were developing a global model for the supply chain process that could be easily adapted through data-driven functionality. The system provided a scalable and generic model for the businesses of various sizes and complexity to utilize data driven analytics.

We are also utilizing data science for customer relationship management (CRM). CRM applications range from understanding our customer needs, predicting behavior and purchasing patterns, optimal pricing strategies and sales force allocations.

Adapting a Data Science Mindset

Strong support from leadership and cross-functional collaboration among domain experts from business units, data scientists and IT teams are enabling these successful outcomes across the company. To accelerate transformation and to further strengthen collaboration, we have established company-wide data science best practices and are developing a training curriculum to attract, retain and develop top-notch data science talent.



Field data: Monsanto has developed a new scientific approach to analyze data layers in the billions.

Photos courtesy of Monsanto.

With the advancement of digital tools, the responsibilities of our workforce are also evolving. We are preparing for an ag workforce that's ready to leverage the full potential of data science in two ways: 1) driving young talent toward STEM careers in ag through education initiatives, and 2) expanding the skill sets of experienced ag professionals (e.g., specialists in plant breeding and biotechnology).

We see data science, AI and other digital tools as an enabling force for human innovation and productivity toward beneficial outcomes for people and the planet. Our commitment to people development, data stewardship and storage excellence, and collaboration with innovators across sectors, is driving the success of our digital transformation. Over the last three years Monsanto has matured as a data science company, and given the exciting discovery and development happening across technologies, shows no signs of slowing down. We anticipate that our continued growth as a company that operates with data science at its core won't slow down either. **ORMS**

Shrikant Jarugumilli (Shrikant.jarugumilli@monsanto.com) leads the Operations Research team within Products & Engineering at Monsanto. He is Monsanto's representative to the INFORMS Roundtable. The author acknowledges input from Adrian Cartier, Seed Production & Customer Experience Data Science Lead; Barry Surber, Global Supply Chain Quantitative Modeling Technologist; Naveen Singla, Data Science Center of Excellence Lead; and Nathan Vanderkraats, Genomics Data Science Lead.

We see **data science, AI** and other **digital tools** as an **enabling force** for **human innovation** and productivity toward **beneficial outcomes.**

EDITOR'S NOTE

INFORMS Editor's Cut on agribusiness, edited by Robin Lougee and Saurabh Bansal, features research and industry articles, as well as videos on the topic. Check it out at: <https://pubsonline.informs.org/editorscut/agribusiness>.