



# Big Data and Machine Learning: What Is It and Can We Use It for 4R Nutrient Management?

By Leanna Leverich Nigon, Director of Agronomy, The Fertilizer Institute

| July 19, 2023

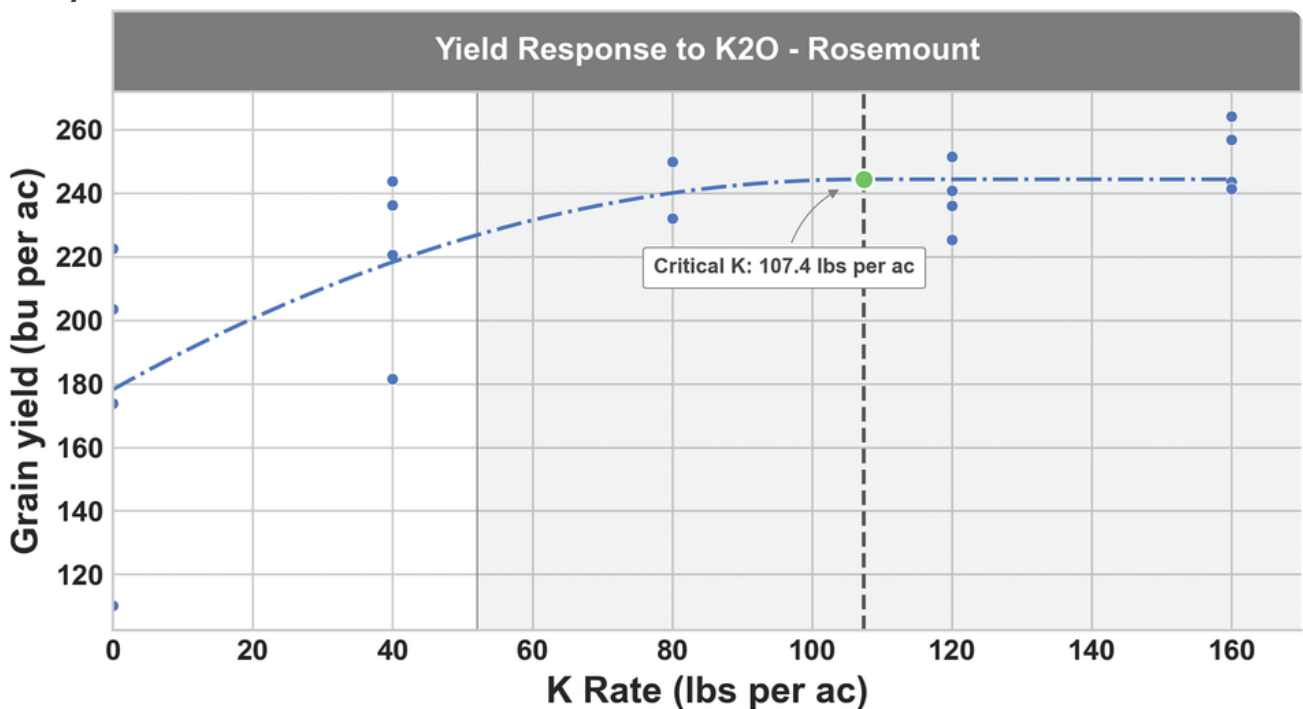


Figure 1. The optimal potassium rate in corn with a quadratic plateau model ( $y = a + bx + cx^2$ ).



Big data and machine learning have the potential to transform agriculture and 4R nutrient management practices.

The integration of these technologies empowers farmers to adapt to variable conditions, optimize applications, and minimize environmental impact. While

challenges such as data quality must be addressed, the future prospects are promising. Earn 1 CEU in Nutrient Management by reading this article and taking the quiz at <https://web.sciencesocieties.org/Learning-Center/Courses>.

---

In recent years, the availability of on-farm data has spurred the expansion of precision agriculture to include big data approaches like machine learning (ML). With proper execution, ML can allow for better implementation of the 4Rs, leading to increased productivity and reduced environmental impacts. However, for growers and crop advisers to properly select and use ML, it is important to understand what modeling is, how it can be used in nutrient management, and its challenges and limitations.

## **Ag Modeling and Machine Learning**

When vetting commercially available ag modeling services, it is important to understand the buzzwords, particularly in the context of agronomy. Firstly, “ag modeling” is already common practice in agricultural sciences with both traditional statistical models as well as more modern ML models being used by researchers.

Traditional statistical modeling uses equations to characterize the relationship between variables, such as the relationship between yield and potassium rate, as depicted in Figure 1. Figure 1 shows the optimal potassium rate in corn with a quadratic plateau model ( $y = a + bx + cx^2$ ). Traditional statistical modeling is commonly used by University Extension to determine crop response to fertilizer rate and develop fertilizer recommendations.

Similar to traditional models, ML models also have their own equations. The ML equations can be complex or simple. Where ML stands apart from traditional modeling is that the computer “iterates” to optimize the model fit to the data in a process that includes “training,” “tuning,” and “testing.” During the many iterations, the computer “learns” the optimal parameters of the equation. For a simple example, an ML algorithm could be developed to optimize the equation values  $a$ ,  $b$ , and  $c$  in the model in Figure 1 (the slope, the intercept, and the quadratic term [the apex of the parabola]). Through the iterations, the computer can find patterns in the data that may not be evident to a researcher to optimize a model.

When to use traditional statistical models versus ML models mostly depends on the agronomic objective. Traditional statistical approaches use an experiment design, like a plot or strip trial, that is randomized and replicated and evaluates a “treatment” such as nitrogen rate as it impacts yield.

In contrast, ML can use datasets with many interacting treatments, non-traditional experimental designs, and various spatial scales. Simply put, it does not require the experimental setup like traditional statistics does.

Machine learning statistics do not rely on the classical statistical power from randomization and replication. Instead, ML statistics rely on cross-validation. Cross-validation is a process where the dataset is split into two groups: data to train the model and data to test the model. The model error in the test dataset indicates how well the model performs (i.e., accuracy of predictions).

Machine learning models are best suited for understanding a physical phenomenon from various measurements that could be interacting with one another. For example, crop yield could be predicted with an ML model using a suite of field factors including weather, historical yield data, and soil information. There is no specific “treatment” impacting yield that we are trying to measure, but rather a larger physical phenomenon with multiple variables impacting it.

When it comes to nutrient management, ML algorithms are built to predict crop nutrient status, optimal nitrogen (N) rates, and growth stage or to create management zones. This is largely possible because of the amount of data available from each farm.

## **Big Data and Data Sources**



In general, farms make great candidates for ML because of the immense amount of data that can be collected each year, making them sources of “big data.” Many growers have been collecting yield data for more than a decade. In addition to yield data, growers may have soil test data and valuable management information (crop, rotation, planting date, etc.). All this on-farm data along with environmental data can be powerful in explaining yield variability and holds great promise for building ML models (Bullock, 2019).

*In general, farms make great candidates for ML because of the immense amount of data that can be collected each year, making them sources of “big data.” Photo courtesy of Adobe Stock/Monopoly919.*

In addition to on-farm management data, a plethora of sensors have been released in the ag marketplace to offer more opportunity for data collection. The 2022 Precision Agriculture Dealership survey shows that more than half (55%) of retailers offer UAV or drone imagery and 18% offer chlorophyll or greenness sensing. These and other sensors, like weather station sensors, can provide vast amounts of data to describe soil conditions, weather patterns, and field variability that can be used in ML models. However, in considering what sensors and measurements are necessary, growers and consultants should always keep in mind the question they are aiming to answer.

If the goal is to optimize N applications, the first step is to identify the field features that influence N. Field factors like organic matter or a greenness measurement from a near-infrared (NIR) camera sensor have a strong relationship with N dynamics. Therefore, these field features would likely be important to measure and include in a ML model to predict N status.

Throwing as much sensor data as possible into an ML model is not the answer to more accurate nutrient predictions/recommendations. Using unnecessary data can actually

create “noise” in any model and cause inaccurate predictions. It could be a red flag if companies are throwing the kitchen sink at problems as they might be lacking the foundational agronomic knowledge. In choosing what sensors to use and data to collect, lean into agronomic knowledge to inform which measurements will be most valuable. Machine learning models in conjunction with agronomic knowledge are the best path forward to achieve accurate predictions (Chlingaryan et al., 2018).



*The 2022 Precision Agriculture Dealership survey shows that more than half (55%) of retailers offer UAV or drone imagery. Photo courtesy of Adobe Stock/america\_stock.*

## **Applications for Machine Learning in 4R Nutrient Management**

In regard to 4R nutrient management, interest has grown in using ML to optimize the right rate, right timing, and right placement of N in particular.

According to the 2022 Precision Agriculture Dealership Survey from Purdue, 49% of growers are using variable-rate technologies (VRT), and just short of 90% of dealers offer VRT services for nutrient applications (Erickson, 2022). It is common to use VRT for site-specific applications of P and K based on soil test levels, but N is more of a

challenge. Unlike the other two primary nutrients (P and K), N is highly reactive in soils and often changes forms (e.g.,  $\text{NH}_4$  to  $\text{NO}_3$  through nitrification). The transformation of N in soils is often influenced by moisture and temperature, making it difficult to predict the optimal N rate for any area of the field.

Given the challenge of optimizing N rates, ML solutions have been pursued in the hopes that the computer may be able to find data patterns to make better N recommendations. Some ML approaches include: (1) predicting crop N status, (2) predicting the optimal N rate, (3) forecasting crop growth, and (4) creating management zones for N applications.

The first approach uses data from sensors and satellite imagery along with additional field information to build ML models that can identify areas within fields with nutrient deficiencies. Farmers can then tailor nutrient application to places in the field where fertilizer is needed, optimizing the **right place** of the 4Rs. However, this strategy requires substantial collection of remotely sensed data and the ability to apply nutrients throughout the season. High-value crops have been an entry point for this type of technology.

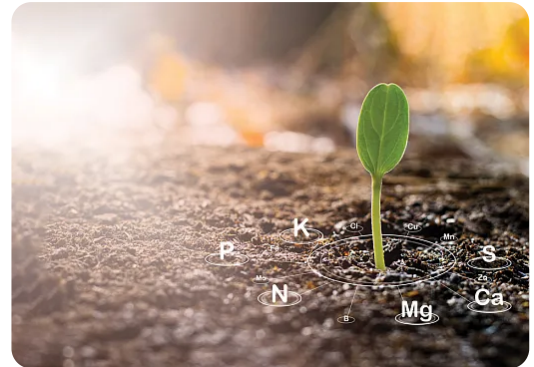
A second ML approach used for nutrient management is predicting the economically optimal N rate (EONR) within fields or for a whole field, the **right rate** of the 4Rs. Researchers have found mixed success with this approach. A 2022 study in of canola production in Canada found that using ML models could reasonably predict optimal N rates for split applications when historical and current weather conditions were included in the model (Wen et al., 2022). However, a Nebraska study in corn found that ML models built from data in one field could not be used to accurately predict N rates in other fields (de Lara et al., 2023). There is much to learn and improve in this area of

ML prediction. Researchers have concluded that as technology improves, there is great opportunity to employ ML, but today more research is required to widely deploy it for N rate predictions (de Lara et al., 2023; Chlingaryan et al., 2018).

Machine learning models can also be built to forecast crop growth stages throughout the season. Knowing the growth stages of the crop can allow for tailoring fertilizer applications to the time of greatest nutrient demand. In this manner, the **right time** of the 4Rs can be improved (Yue et al., 2020).

Other management strategies such as the creation of management zones can also be aided by ML models. Using data from various sources, including topography maps, weather stations, and soil data, ML algorithms can be used for zone creation and to identify the most important field features for guiding nutrient prescriptions. With zones established, nutrient applications can be customized for each zone (Jaynes et al., 2011; Nawar et al., 2017).

A better understanding of nutrient status and crop yield across fields can help increase the accuracy of nutrient applications, minimizing unwanted environmental impacts and costs for growers. Predicted nutrient status, optimal N rates, crop growth stage, and delineation management zones can all be used to optimize the 4Rs, ensuring that nutrients are supplied at the right place, time, and right rate.



*A better understanding of nutrient status and crop yield across fields can help increase the accuracy of nutrient applications, minimizing unwanted environmental impacts and costs for growers. Photo courtesy of Adobe Stock/Gamogamo.*

## **Challenges in Machine Learning and Ag Modeling**

## Data Quality and Farm Records

While the integration of big data and ML in agriculture offers immense potential, several challenges must be addressed. Data quality is particularly challenging in agriculture, and the old adage “garbage in, garbage out” applies to ML efforts in nutrient management (Nielsen, 2020). Farm data full of error leads to poor-performing ML models and inaccurate predictions. Good record keeping (planting dates, crop rotations, etc.) is

essential to train ML models as this information is their backbone. Historic weather data or satellite imagery can be gathered from public sources going back several years, but the only source of farm information is from the grower. Therefore, keeping good farm records is imperative for employing this type of technology. Other data cleaning like yield monitor calibration should also be completed to ensure predictive models built from on-farm data will be as accurate as possible.

## Data Privacy

Privacy and data ownership are another significant consideration. Farmers must have confidence that their data is protected and that they retain control over its use and dissemination. On-farm data provided by the grower is necessary to build ML models. While the ML models created by private companies using that data may be proprietary, the proprietary status of the on-farm data has yet to be addressed by the industry and policymakers. Collaborative efforts among farmers, technology providers, and policymakers are necessary to develop guidelines and frameworks that safeguard



*Crop yield could be predicted with a machine-learning model using a suite of field factors including weather, historical yield data, and soil information. Photo by Tao Wang.*

data rights, privacy, and ownership for ML modeling to flourish in the ag industry.

### **Lack of User-Friendly Software**

The development of user-friendly interfaces, farm databases, and decision support systems will also be crucial to ensure that farmers can easily interpret and apply the insights derived from big data and ML technologies. Furthermore, the industry has hurdles with software compatibility. Developing software that is compatible and can communicate with existing systems that farmers already have in place will be necessary to increase adoption of ML technology. Education and training programs will also play a pivotal role in equipping agronomists and farmers with the necessary skills to leverage these tools effectively.

### **Future of Machine Learning in Nutrient Management**

Big data and ML have the potential to transform agriculture and 4R nutrient management practices. The integration of these technologies empowers farmers to adapt to variable conditions, optimize applications, and minimize environmental impact. While challenges such as data quality must be addressed, the future prospects are promising. Growers and agronomists can prepare for the use of ML by being diligent in organizing and recording their farm practices. Farm management information will be key to help build these models.

Continued advancements in ML and increased collaboration among stakeholders will drive the adoption of ag tech in agriculture, providing a meaningful tool for implementing sustainable and efficient 4R nutrient management practices.

### **References**

Bullock, D.S., Boerngen, M., Tao, H., Maxwell, B., Luck, J.D., Shiratsuchi, L., Puntel, L., & Martin, N.F. (2019). The data-intensive farm management project: changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, 111, 2736–2746. <https://doi.org/10.2134/agronj2019.03.0165>

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>

de Lara, A., Mieno, T., Luck, J.D., & Puntel, L. (2023). Predicting site-specific economic optimal nitrogen rate using machine learning methods and on-farm precision experimentation. *Precision Agriculture*. <https://doi.org/10.1007/s11119-023-00181-8>

Erickson, B., & Lowenberg-DeBoer, J. (2022). 2022 Precision agriculture dealership survey. Department of Agricultural Economics and Agronomy, Purdue University & CropLife Magazine. [https://ag.purdue.edu/digitalag/\\_media/croplife-report-2022.pdf](https://ag.purdue.edu/digitalag/_media/croplife-report-2022.pdf)

Jaynes, D.B., Kaspar, T.C., & Colvin, T.S. (2011). Economically optimal nitrogen rates of corn: Management zones delineated from soil and terrain attributes. *Agronomy Journal*, 103(4), 1026–1035. <https://doi.org/10.2134/agronj2010.0472>

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A.M. (2017). Delineation of soil management zones for variable-rate fertilization. *Advances in Agronomy*, 143, 175–245

Nielsen, R.L. (2020). Yield monitor calibration: garbage in, garbage out. Corny News Network, Purdue University Agronomy Extension. <http://www.kingcorn.org/news/timeless/YldMonCalibr.html>

Wen, C., Ma, B., Vanasse, A., Caldwell, C.D., & Smith, D.L. (2022). Optimizing machine learning-based site-specific nitrogen application recommendations for canola production. *Field Crops Research*, 288(4), 10877. <https://doi.org/10.1016/j.fcr.2022.108707>

Yue, Y., Li, J.H., Fan, L.F., Zhang, L.L., Zhao, P.F., Zhou, Q., ... & Dong, X.H. (2020). Prediction of maize growth stages based on deep learning. *Computers and Electronics in Agriculture*, 172, 105351. <https://doi.org/10.1016/j.compag.2020.105351>

### Self-Study CEU Quiz

Earn 1 CEU in Nutrient Management by taking the quiz for the article at <https://web.sciencesocieties.org/Learning-Center/Courses>. For your convenience, the quiz is printed below. The CEU can be purchased individually, or you can access as part of your Online Classroom Subscription.

1. **In machine learning, the algorithm “learns” how to fit models by**
  - a. iterating through the data.
  - b. searching the Web.
  - c. mining additional datasets.
  - d. using randomization and replication.
2. **Machine-learning modeling differs from traditional statistical modeling by**
  - a. using complex equations.
  - b. not requiring standard experimental design.

- c. relying on randomization and replication instead of cross-validation.
- d. being based on assumptions instead of being data-driven.

**3. Error is quantified in machine learning models using**

- a. standard deviation.
- b. replication and randomization.
- c. cross-validation.
- d. the experimental treatment.

**4. Farms considered great candidates for ML in agriculture because**

- a. of their small size and limited data collection capabilities.
- b. ML models work best with limited datasets.
- c. farms produce high quality sensor data.
- d. farms generate large amounts of data (big data).

**5. In 2022, \_\_\_\_% of growers were using variable-rate technologies (VRT) for nutrient applications.**

- a. 65
- b. 55
- c. 49
- d. 33

**6. What are some challenges in implementing ML in nutrient management?**

- a. Not enough sensors available on the marketplace to collect data.
- b. Limited computing power for ML algorithms.
- c. Data quality and availability of farm management data/records.
- d. Low interest from growers and agronomists.

**7. In nutrient management, ML algorithms can be built to predict**

- a. optimal nitrogen (N) rate.

- b. crop nutrient status.
- c. crop growth stage.
- d. All of the above

8. **Nitrogen is considered a reactive nutrient in the soil. Nitrogen transformation in soils is influenced by \_\_\_\_\_ and \_\_\_\_\_.**

- a. temperature; phosphorus level
- b. temperature; moisture
- c. organic matter percentage; electrical conductivity
- d. moisture; soil bulk density

9. **Which of the following statements is true about ML models and data quality?**

- a. High quality data leads to better-performing ML models and accurate predictions.
- b. Machine-learning (ML) models can compensate for poor data quality, so it is not a significant concern.
- c. Machine-learning (ML) models rely solely on public sources of data, not on-farm data.
- d. Machine-learning (ML) models do not require historical data, only real-time measurements.

10. **Growers and agronomists interested in using ML now or in the future should prepare by**

- a. buying as many sensors as possible to collect data.
- b. organizing farm records from years past and keeping good records for the future.
- c. not wasting time calibrating yield monitors or other equipment.

d. cleaning out and deleting historical farm records.

---

*Text © . The authors. CC BY-NC-ND 4.0. Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.*