On-Farm Evaluations to Calibrate Tools for Estimating Late-Season Nitrogen Status of Corn

P. M. Kyveryga* and T. M. Blackmer

ABSTRACT

Properly calibrated diagnostic tools are needed to evaluate the performance of different N management practices for corn (Zea mays L.). Until now, mostly controlled studies were used for such calibrations. We utilized on-farm evaluation studies to verify current and identify new N status categories using the corn stalk nitrate test (CSNT) and late-season digital aerial imagery of the corn canopy. From 2007 through 2010, producers conducted 125 trials across Iowa. Each trial had treatments of a producer’s normal N rate alternated with a rate that was about one-third lower or higher. Categorical yield response (YR), expressed as profitable and unprofitable, was related to green reflectance (GR), relative green reflectance (RGR), or CSNT sampled within nine areas in each trial. Multilevel binary logistic regressions were used to estimate the probability of receiving profitable YR for a range of CSNT, RGR, and GR values. Among the three diagnostics, RGR performed slightly better but required applying at least two N rates within producers’ fields. For CSNT, the identified optimal category was almost the same as that currently recommended in Iowa (700–2000 mg NO3–N kg–1), even when corn and N prices deviated from their long-term averages by 30%. Due to the uncertainty in N availability, however, the critical CSNT value for fall manure treatments was about 3000 mg NO3–N kg–1 higher than that for fall anhydrous NH3, spring urea–NH4NO3, or sidedress N. On-farm studies can be used to calibrate late-season N diagnostic tools for evaluating management practices that differ in rates, forms, and timing of N applications.

During the last decade, corn production in the U.S. Midwest has been under increasing economic, environmental, and regulatory pressures. As a result, the general public, producers, and agronomists are looking for technologies and diagnostic tools that can substantially optimize N fertilizer management and reduce economic and environmental risks by increasing fertilizer use efficiency and minimizing N losses to water and the air.

Historically, N fertilizer recommendations for corn in the U.S. Midwest have been focused only on prescriptive strategies by making general recommendations based on data collected in controlled, usually small-plot, N yield response studies (Sawyer et al., 2006; Vanotti and Bundy, 1994). The objective of such studies is to estimate economic optimum N rates (EONRs), rates that maximize returns to N after the fact, and to extrapolate these rates to other geographic areas and future growing seasons. Because of large spatial variability in rainfall, soil properties, and yield response to N, after-the-fact estimates of EONRs can carry a large degree of uncertainty (Scharf et al., 2005). To reduce this uncertainty, N yield response studies are often conducted under on-farm conditions to better represent management practices used by producers, which include tillage, plant genetics, fertilizer application equipment, N fertilizer form, and timing of application.

On-farm studies are also important for regulation and monitoring purposes. For example, producers may be required not only to use the best management practices and current N fertilizer recommendations but also to document that the management practices used are efficient, environmentally friendly, and sustainable (USEPA, 2011). To provide these site-specific assessments, producers will need to evaluate N management at both farm and field levels (NRCS, 2011). As a result, on-farm evaluations are becoming crucial for providing feedback on corn N status and collecting data that can be used to improve and refine current N fertilizer recommendations and improve producers’ profitability (Blackmer and Kyveryga, 2010; Tao et al., 2012).

The late-season CSNT can be used to document how N management practices perform at the field scale. The test is a diagnostic tool that can identify whether corn plants had enough N, too little, or more than needed in after-the-fact assessments. The test was originally developed in Iowa (Binford et al., 1990, 1992) and proved to be reliable in other studies across the country (Brouder et al., 2000; Fox et al., 2001; Wilhelm et al., 2005). Iowa uses four N sufficiency categories: deficient, marginal, optimal, and excessive (Blackmer and Mallarino, 1996); other states use only three categories: low, optimal, and excessive (Beegle and Rotz, 2009; Brouder et al., 2000; Hooker and Morris, 1999).

The reliability of late-season N evaluations partially depends on whether diagnostic tests are properly calibrated and whether the tests correctly classify observed stalk NO3 values into the

Abbreviations: AA, anhydrous ammonia; AIC, Akaike information criterion; CSNT, corn stalk nitrate test; GPS, global positioning system; GR, green reflectance; MBLR, multilevel binary logistic regression; RGR, relative green reflectance; UAN, urea–ammonium nitrate; YR, yield response; C–S, corn after soybean; C–C, corn after corn.
defined N sufficiency categories. The test calibration usually requires relating stalk NO₃ values to YR or EONR values. The assumption is that corn plants that test in the optimal category received N equal to the EONR, plants testing deficient received less N than the EONR, and plants testing excessive received more N than the EONR. However, the EONRs estimated from controlled N response trials are often estimated with large variability and uncertainty (Hernandez and Mulla, 2008; Jaynes, 2011), and conducting multivariate yield response trials under on-farm conditions is not always practical. Thus, alternative methods are needed to verify the original CSNT calibration categories and, if needed, to identify new categories based on on-farm evaluations.

The original CSNT calibration studies in Iowa were conducted using one N source [NH₄(SO₄)] applied to corn in spring before planting (Binford et al., 1990, 1992); however, current N management practices include the use of at least four different N forms or sources (urea–NH₄NO₃ [UAN], anhydrous NH₃ [AA], urea, or animal manure), three different timings for N application (fall, spring [preplant], and sidedress), and three methods of application (injected, broadcast on the soil surface, or banded). The current stalk N sufficiency categories were also identified when corn and N fertilizer prices were much lower than they are currently. The question becomes whether the current CSNT categories are appropriate or need adjustments for the evolution in N management practices over time in Iowa and changes in the economics of N fertilization.

Another tool that can be used for late-season evaluations is digital aerial imagery of the corn canopy. Digital aerial imagery can be used in different ways. For example, a large study conducted across Iowa used digital aerial imagery to guide CSNT sampling within about 1500 corn fields during 2 yr to identify differences between N management categories, which were formed based on a combination of N source and the timing of application (Kyveryga et al., 2010). Also, late-season digital aerial imagery was used to estimate the percentage of area with sufficient and deficient corn N status within about 1600 corn fields during 3 yr (Kyveryga et al., 2011). Using digital aerial imagery directly to evaluate late-season N status, however, is complicated by a lack of an appropriate calibration methodology for on-farm studies.

Precision agriculture technologies such as remote sensing and yield monitoring are well suited for on-farm evaluation studies (Griffin et al., 2008). For a relatively low cost, data collected in such studies can potentially be used to answer an array of important management questions. For example, farmers are often interested not only in whether a given corn field had enough N but also in whether they could correct an N deficiency by applying additional N, usually about 30 to 60 kg N ha⁻¹. Another common question is whether the CSNT test, in addition to field-average corn N status, can provide some site-specific N assessments within spatially variable fields.

The main objective of this study was to develop a method for calibrating late-season N diagnostic tools using CSNT and digital aerial imagery of the corn canopy to identify corn N status using YR observations collected from on-farm evaluation trials conducted at multiple locations across Iowa during 4 yr. In addition, we wanted to demonstrate how on-farm observations, which included different N management practices (forms and timing of applications) with different levels of applied N, can be used to verify the existing categories of N status and, if needed, to identify new categories for late-season N-status evaluations.

**MATERIALS AND METHODS**

**Field Methodology and Yield Data Collection**

One-hundred twenty-five on-farm evaluation trials were conducted during four growing seasons. Forty-two trials were conducted in 2007, 20 trials in 2008, 34 trials in 2009, and 29 trials in 2010. The trials were located across Iowa, with the majority located in the central and north-central parts of the state (Fig. 1).

There were three categories of trials: “normal minus 50” and “normal plus 50” for commercial N sources and “normal plus 50” for injected liquid swine manure. In the first two categories, producers used their normal N rates and compared them with rates that were either 30% (or 56 kg N ha⁻¹) lower or 30% higher than the normal. For the injected liquid swine manure, producers compared their normal N manure rates, based on total N, with rates that were 30% (or 56 kg N ha⁻¹) higher. Normal rates are defined here as the average N rates used by producers during the last several years. The 56-kg increment was chosen because it is the common amount of N used for early spring (preplant) applications, with or without herbicides, or for in-season supplemental or rescue N applications. In addition, fields fertilized at the near-optimal N range would often need about 56 kg N ha⁻¹ to produce a detectable YR.
The two N treatments were labeled as high and low, with the high rate being the "normal rate plus 50" or "manure normal rate plus 50," and the low rate being the "manure normal rate" or a "normal rate minus 50." The amount of N applied for a "normal N rate" included the total amount of N applied from all N sources and all times of application during the growing season. The manure N rates (based on total organic and mineral N) were estimated as averages of the N contents of three or four manure samples collected by producers during the application. In about 20% of the trials, the manure N rates were estimated from manure analysis done before or after manure applications for the same manure source from the same livestock facility. In about 10% of the trials, the N content in the manure was estimated based on book values.

The two N treatments, low and high, were applied in strips replicated at least three times within each trial (Fig. 2). Some of the trials had more than five treatment replications. The width of the strips varied depending on the width of the fertilizer application equipment used and on the width of the combine header. The strips were positioned to capture the maximum range of variability in soil topography and soil organic matter within the fields. The strips went the full length of the fields and covered at least 10 ha. The N treatments were applied in alternating strips to avoid application errors and facilitate a common way to calculate yield differences between the two treatments for all trials. In about 80% of the trials, the strip locations were recorded by an onboard global positioning system (GPS) during the applications. In the rest of the trials, farmers flagged the strips during the treatment applications and then recorded GPS coordinates for each treatment.

![Fig. 2. Sampling areas (5–12 by 30 m) for collecting corn stalk nitrate test samples and extracting green reflectance values of the digital aerial imagery of the corn canopy. Each trial had two N rates: normal (defined as high or low) and normal minus 56 kg N ha⁻¹ (defined as low) or normal plus 56 kg N ha⁻¹ (defined as high). The lower rate is shown as the lighter strips and the higher rate is shown as the darker strips on the image. The sampling areas were selected to represent the wide range of variability in reflectance of the corn canopy within each trial.](image)

![Table 1. Average total N rates applied by producers to normal N treatments for different N management categories for corn after soybean (C–S) and corn after corn (C–C) in 125 on-farm evaluation studies conducted between 2007 and 2010 across Iowa.](table)
relatively low probability of a profitable YR would be considered N sufficient; likewise, those that show a relatively high probability of a profitable YR would be considered N deficient.

Yield responses were classified into two binary categories: profitable and unprofitable. For identifying the binary categories, we used long-term N fertilizer (US$1.01 kg\(^{-1}\)) and corn (US$200 Mg\(^{-1}\)) prices, at which 0.31 Mg of corn would cover the cost of the additional 56 kg of N.

Collection of Digital Aerial Imagery

Digital aerial imagery of the corn canopy was collected in late August or early September when corn N stress is easily detectable because corn plants often deplete the N supplied by fertilizer and by the soil. Each aircraft used four 12-bit digital cameras with a charge coupled device (CCD) array of 1600 by 1200. The imagery was taken during cloud-free days from a height of about 2400 m above the ground surface. The imagery had four bands, with the blue band capturing a spectral range from 410 to 490 nm, the green band from 510 to 590 nm, the red band from 610 to 690 nm, and the near-infrared band capturing from 800 to 900 nm. Twenty to thirty individual images were taken for each field and then ortho-mosaicked into one composed digital image for the entire field. The final composed digital images were geographic information system (GIS) ready, geo-referenced, and tonally balanced, with a spatial resolution of about 1 m. The images were ortho-rectified using USGS 7.5-min digital elevation models. The composed 12-bit images were converted to eight-bit data to reduce the amount of computer memory used in displaying and processing the imagery.

The composed, uncalibrated imagery was enhanced by extending the dynamic range, the range of digital reflectance values between the darkest and the lightest parts of the imagery, within a field, for each band using ERDAS Imagine software. The enhanced imagery had about 80% of reflectance values between ±2 SD from the mean reflectance for each band. The average dynamic range was from about 50 to 80 digital counts across all of the fields. The enhanced imagery showed better visual differences in the corn canopy reflectance than the raw, unenhanced imagery and better captured the potential spatial variability in corn N status. Unlike satellite imagery, which is radiometrically corrected to the absolute reflectance values, our digital aerial imagery for each field was not radiometrically corrected because it was taken by different cameras and at different times in each year. The enhancement procedure, however, partially normalized the reflectance values of the imagery taken by the same digital cameras across a large number of other corn fields surveyed without N fertilizer treatments (Kyvernyga et al., 2012).

The enhanced color imagery (blue, red, and green bands) was overlaid with N treatment maps to select nine sampling areas to collect CSNT samples from each treatment—three for each pair of two N treatments within each field (Fig. 2). The areas were selected to represent a wide range of variability in the corn canopy within each field. Coordinates for the selected areas were recorded and uploaded to handheld GPS units. Agronomists and crop consultants used GPS units to navigate to the sampling locations within each field. The sampling locations were recorded again as the samples were collected to verify the exact coordinates.

Spatially interpolated (4-km grid) monthly rainfall data for Iowa were downloaded from the Iowa Environmental Mesonet (http://mesonet.agron.iastate.edu/rainfall/). Each trial was assigned one rainfall value for each month from the nearest rainfall grid using the Spatial Join function of ArcGIS Desktop 9.3.1 software (Environmental Systems Research Institute).

Collection of Corn Stalk Nitrate Samples

Fields were sampled after the corn plants reached physiological maturity (black-layer stage) according to the original CSNT guidelines (Blackmer and Mallarino, 1996). In each sampling area, 10 individual cornstalk segments of 10 corn plants were collected from within two 15-m segments of a corn row. Barren, damaged plants and plants in areas of abnormally low or high plant density were not sampled. The samples were cut between 15 and 35 cm above the ground to produce 20-cm-long stalk segments. The stalk samples were shipped to the laboratory usually on the same day or always within 2 d after sampling.

The stalk samples were dried at 65°C and ground to pass a 1-mm mesh. The samples were extracted with 2 mol L\(^{-1}\) KCl and the extracted solutions were filtered and analyzed for NO\(_3\)–N values with a Lachat flow injection analyzer (Lachat Instruments).

Analysis of Digital Aerial Imagery

To extract reflectance values for each band, polygons about 30 m long and about half of the treatment width (about 5–12 m) and positioned in the center of each strip were drawn for each stalk sampling area in each trial. There were nine sampling polygons for each of the two N treatments in each field (Fig. 2). Reflectance values from the digital aerial imagery were extracted by building a GIS model using the Zonal Statistic and Append tools of Spatial Analyst for ArcGIS. Summary reflectance values (i.e., mean, median, SD, and a number of pixels) of the four bands were extracted and combined into one summary table. Mean reflectance values were used to calculate the RGR, a ratio of green reflectance at the low N rate to green reflectance at the high N rate. Relative green reflectance is an index of response to N, with values <1 indicating no response or negative response and values >1 indicating positive response. Several vegetative indices, such as the normalized difference vegetative index and the green normalized difference vegetative index, were also calculated using the near-infrared, red, and green band reflectance values. These indices did not provide advantages for identifying N sufficiency categories when compared with the green band alone or the RGR.

To calculate yield differences between the two treatments that would correspond to stalk NO\(_3\) sampling areas, polygons of about 50 m long and the full width of the treatment were used to extract mean yields in the stalk sampling areas. The yields were extracted using the Intersect function of the ArcGIS software. Yield responses were calculated as differences between the yield at the high N rate and yield at the low N rate in each of the nine areas sampled for the CSNT.

Statistical Analysis

Plant or tissue test calibration procedures in soil fertility studies require pooling YR observations from trials conducted at many locations during several years to capture the wide range of variability in soil properties and weather and to address potential differences between corn hybrids and various N management practices. Because the digital aerial imagery in our study was not radiometrically corrected and correcting
the imagery to reference satellite imagery would be time consuming and expensive, pooling YR data across many fields would ignore differences in corn N status between individual fields. To avoid the problem of complete pooling of the data, we utilized multilevel regression analysis. The multilevel (also called hierarchical or random effects) model allows simultaneous modeling of different sources of random variability observed at different scales or levels (Cressie et al., 2009; Gelman and Hill, 2007; McMahon and Diez, 2007). Multilevel models belong to the general class of generalized linear models in which model parameters vary at more than one level. The basic idea is in partial pooling, a compromise between complete pooling (complete aggregation) by ignoring differences within fields and complete independence (complete disaggregation), which models relationships separately in each field by ignoring differences among the fields. The partial pooling effect is accomplished by modeling the relationship simultaneously at two levels. For the first level, separate regressions are fit to data for each field, addressing within-field variability. For the second level, regression parameters (e.g., slopes and intercepts) from individual fields are expressed as a separate random model with its own distribution parameters such as mean and SD. The parameters for the second level model characterize both among- and within-field variability. Unlike ordinary regression analysis, multilevel analysis makes corrections for possible biases in parameter estimates resulting from subsampling or clustering within fields, and it assumes that individual observations within the same field correlate or are more similar than those observed in other fields (Guo and Zhao, 2000).

We used multilevel binary logistic regressions (MBLR) in which the binary response variable (profitable or unprofitable YR) was regressed against a single predictor such as GR, RGR, or CSNT values measured in each of the 125 trials conducted during 4 yr. The multilevel binary logistic regression allowed estimation of the probability of a profiable YR across a range of values of the independent variables. We used the logit link function, \( \log(\Pr(1 - P_{ij})) \), which transformed continuous values of the independent variables to a range of predicted probabilities from 0 (unprofitable) to 1 (profitable).

The MBLR had two levels: within field and field. For the within-field level, separate regressions were fit to individual observations in each trial, and intercepts and slopes of binary logistic regressions varied among the trials. For the field level, intercept and slope were estimated from the parameters of the within-field models from individual trials. The intercept and slope for the field-level model did not vary and were considered as fixed, while those for the within-field-level model could vary and were considered as either fixed or random.

A general multilevel binary logistic model for predicting a binary economic YR, \( y_{ij} \), at two levels with one predictor, \( x_{ij} \), measured at the \( i \)th sampling location within the \( j \)th trial can be described for a within-field-level model as

\[
\Pr(y = 1) = \log \left( \frac{p_{ij}}{1 - p_{ij}} \right) = \alpha_j + \beta_j x_{ij} \tag{1}
\]

and for a field-level model as

\[
\alpha_j \text{ and } \beta_j \sim N(\mu, \delta^2) \text{ for } j = 1, \ldots, 125 \text{ trials} \tag{2}
\]

where \( p_{ij} \) is the probability of profitable YR; \( \alpha_j \) is a random intercept and \( \beta_j \) is a random slope, estimated separately for each trial; \( \mu \) and \( \delta \) are mean and variance for distributions of the estimated random intercepts and slopes of 125 trials.

To determine whether to use a fixed or random intercept and slope for the within-field model, we compared full and reduced models. The full model included both intercept and slope as random effects; the reduced model included the intercept as random and the slope as a fixed effect. The trial location effect was considered random. The effect of a year was not considered in the analysis because the data consisted of observations from only 4 yr and because of the complexity in calculating parameters for a model with a higher number of levels. The best final MBLR was chosen by comparing Akaike information criterion (AIC) values, with the lower AIC indicating the better and more parsimonious model. When AIC values were the same, a \( \chi^2 \) test with one degree of freedom was performed to determine whether the AIC values for the full and reduced models were statistically different. By using MBLR, we also tested the effects of field-level covariates such as average monthly rainfall and normal N rate on the probability of a profitable YR.

Critical values separating profitable and unprofitable YR were estimated for CSNT, RGR, and GR values. Critical values were estimated from the field-level MBLR, which simultaneously considered both within-trial and among-trial variation. A value of 0.51 was considered as a cutoff point separating predicted probabilities into two binary categories of YR. Predicted categories of YR were compared with the observed categories. The percentage of all correctly predicted samples (profitable and unprofitable), percentage sensitivity (profitable), and percentage specificity (unprofitable) were calculated performing cross-tabulation analysis. In addition, the agreement between predicted and observed data was assessed by using the Cohen’s kappa index, which adjusts for a possible random match between the observed and predicted categories (Cohen, 1960).

Parameters for multilevel models were estimated using the “lme 4” package (Bates and Maechler, 2010), and kappa index values were estimated using the “concord” package (Lemon and Fellows, 2009) of the R Statistical Program (R Development Core Team, 2009). Predicted probabilities for the field-level models were estimated using the inverse logit function (invlogit) of the “arm” package (Gelman and Hill, 2007).

**RESULTS AND DISCUSSION**

**Performance of Corn Stalk Nitrate Test**

Figure 3 shows the relationship between stalk NO₃ values for the low N rate and YR to the additional N (56 kg N ha⁻¹) in each year of the study and for the pooled data across 125 on-farm trials conducted during 4 yr. A relatively large YR in the lower range of stalk NO₃ values, from 0 to 500 mg NO₃-N kg⁻¹, rapidly decreased to a relatively small YR in a range of stalk NO₃ values >1000 mg NO₃–N kg⁻¹. This trend, called luxury N uptake, was observed in the original small-plot studies in Iowa to calibrate the CSNT with a wide range of N rates (Binford et al., 1990, 1992). The relationships between YR and stalk NO₃ values were similar...
in each year (Fig. 3), suggesting that on-farm trials with only two N rates can potentially be used to evaluate the original calibration categories for the CSNT.

Considering that the magnitude and frequency of a YR indicate the level of corn sufficiency in N, relatively large YR values suggest below-optimal or deficient N supply and relatively small YR values suggest an optimal or excessive N supply. Across 4 yr, about 70% of the YR observations in the optimal and excessive ranges were unprofitable because they were located below the line that shows the cost of additional N expressed in amount of grain in Fig. 3E. Conversely, about 60% of the YR observations in the deficient range were profitable. About 20% of the stalk sampling areas, however, had negative YR values in the deficient stalk test category, indicating that the CSNT incorrectly predicted some positive and profitable YR values.

The relationship shown in Fig. 3E is slightly different from that observed in the original calibration studies in Iowa (Binford et al., 1990, 1992), where a linear plateau model was fitted to describe how relative yield (also an index of YR and N sufficiency) changed with an increase in CSNT values. In our study, the relationship was also nonlinear but with a larger degree of heterogeneous variances and variability. The large variability in YR is partially because farmers used normal N rates in the near-optimal range (no zero or extremely low N rates), producing relatively low YR values. For example, about 25% of all sampling areas had YR < 0 Mg ha\(^{-1}\); 20% had YR of 0 to 0.30 Mg ha\(^{-1}\); 30% had YR of 0.30 to 1 Mg ha\(^{-1}\); 20% had YR of 1 to 2 Mg ha\(^{-1}\), and about 5% of the areas had YR > 2 Mg ha\(^{-1}\). In addition, the large variability in YR is partially because yield monitor observations were used in YR calculations and the stalk sampling areas were selected to represent a wide range of spatial variability in the corn canopy color within the trials. In addition to N fertilization, stalk NO\(_3\) values in farmers’ fields could be potentially influenced by other factors such as the amount of rainfall or management. Although the relationship in Fig. 3E suggests that stalk NO\(_3\) values cannot predict YR with a high degree of certainty, the CSNT can easily identify two categories, one with relatively high probabilities and the other with relatively low probabilities of a profitable YR.

**Green Canopy Reflectance for Predicting Yield Response**

Figure 4 shows the relationship between RGR, a ratio of the green reflectance at the low N rate to that at the high N rate, and YR to the additional N in each year and for the pooled data of 4 yr. Unlike the relationship between CSNT values and YR (Fig. 3), the relationship between YR and RGR was linear (except in 2009), with larger YR values corresponding to larger RGR values, and with coefficients of determination (\(r^2\)) for the linear models ranging from 0.16 to 0.63 (Fig. 4). The highest \(r^2\) was observed in 2010, a relatively wet year, and the smallest was observed in 2009, a relatively dry year (Fig. 5). These differences among the years could be explained by the larger YR values observed in the wetter years.

The RGR values provided quantitative assessments of YR within the fields. The critical RGR separating sufficient from deficient samples can be identified by finding an intersection point between the regression line and the marginal cost of additional N expressed as the amount of grain (Fig. 4). The critical RGR value for the pooled data was 1.05, and the critical values were similar among the years. The regression lines extended below a YR of 0 Mg ha\(^{-1}\) and to the left from the critical RGR, suggesting that a large percentage of negative YR can be explained by the corn canopy reflectance. In fact, about 30% of the areas with a negative YR had smaller reflectance values for the low N rate (the darker color on the imagery) than for the high N rate. This can partially be explained by differences between the two
paired sampling areas in corn plant densities, weed pressure, water movement, soil organic matter, errors due to fertilizer application, and other factors. The ability to explain negative YR values, especially those that had stalk NO₃ values below 250 mg NO₃–N kg⁻¹, indicates that the use of the imagery is superior to the use of the CSNT. To calculate the RGR, however, at least two N rates are needed within each trial.

The use of GR measured in sampling areas with the low N rate (data not shown) revealed a similar linear relationship with YR as shown in Fig. 4. Coefficients of determination, however, were about 50% smaller than those for the relationship between RGR and YR, suggesting that normalizing reflectance values to the reference, high N rate is critical to increase the predictability of a YR based on image spectral properties (Shanahan et al., 2001).

The relationships between RGR and YR (Fig. 4) and GR and YR (data not shown) also showed large variation. This large variation, similar to that shown in Fig. 3, is partially due to errors in estimating the YR, large spatial variability within the fields, and the use of imagery that was not radiometrically corrected to the standard reflectance values across all of the fields. Thus, using the pooled data may potentially lead to large uncertainty in YR estimates because the regression analysis shown in Fig. 4 partially ignores differences in corn N status and corn canopy characteristics between individual fields.

Identifying Critical Values for Late-Season Evaluation Tools

Figure 6 shows the method for identifying critical values for classifying sampling areas into profitable or unprofitable categories using CSNT observations and the RGR or GR values of the corn canopy. Multilevel binary logistic regressions with a single predictor (i.e., CSNT, RGR, or GR) were used to simultaneously model the variability observed within and between the trials. The profitable category was chosen as a reference category in the analysis. Therefore, the vertical axis in Fig. 6 shows the probability of receiving a profitable YR. The observed binary categorical values in YR were plotted along the two horizontal lines, with probability values of either 0 (unprofitable) or 1 (profitable). The regression lines show the predicted probabilities estimated based on fixed parameters (intercepts and slopes) of the field-level models, which were estimated from intercepts that varied and slopes that were constant across 125 on-farm trials conducted during 4 yr. All fixed parameters for the field-level models were statistically significant at the 0.01 probability level (Table 2).

The predicted probability of a profitable YR gradually decreased with the increase in CSNT values (Fig. 6A). A sampling area was considered as unprofitable if the probability
when CSNT values were <250 mg NO₃–N kg⁻¹, the predicted probability of a profitable YR was about 0.6, suggesting a moderate need for additional N. When CSNT values were >2000 mg NO₃–N kg⁻¹, the probabilities were <0.4, suggesting a low need for additional N. A vertical dashed line crossing the 0.5 line separates profitable and unprofitable areas, which would be considered as N deficient and sufficient, respectively. The vertical line crosses the x axis at about 1000 mg NO₃–N kg⁻¹, indicating that the currently used optimal category (700–2000 mg NO₃–N kg⁻¹) for the CSNT is adequate to identify the corn N status. The value of 1000 mg NO₃–N kg⁻¹ is a reference or midpoint for the estimated optimal range. The lower and upper limits for the optimal range in Fig. 6 will depend on how users consider the uncertainty in predicting a profitable YR expressed as probability values.

The model in Fig. 6A correctly predicted 75% of observations for the two binary categories of economic YR. Because the specificity percentage was larger than the sensitivity percentage, the model was slightly better in predicting unprofitable sampling areas than profitable ones. This observation is reasonable because a relatively large variation in YR was observed in the deficient range, <250 mg NO₃–N kg⁻¹ (Fig. 3E). From its maximum value of 1.0, Cohen’s kappa chance-corrected index was 0.50, indicating a satisfactory predictability of CSNT across all the trials (Fig. 6A).

The relationship between the predicted probability of a profitable YR and RGR values across all the trials was described by a steep-sloping S-shaped curve (Fig. 6B). The curve shows that the probabilities rapidly increased with higher RGR values. The model in Fig. 6B correctly predicted 78% of the observations for the two binary categories of economic YR. This prediction was slightly better than that for the stalk NO₃ values shown in Fig. 6A and with a slightly higher Cohen’s kappa index value. Similarly to the CSNT, RGR predicted slightly better unprofitable than profitable YR values.

Unlike the probability curve for the CSNT (Fig. 6A), the steep S-shaped curve for RGR makes it easier to identify a single critical value (Fig. 6B). The critical value separating N-deficient and -sufficient categories for RGR was 1.10. This critical value is in

Table 2. Parameter estimates and standard errors for field-level multi-level binary logistic regressions for identifying the probability of a profitable yield response to N fertilizer using observations of the corn stalk nitrate test (CSNT) and relative green reflectance (RGR) or green reflectance (GR) of the late-season digital aerial imagery of the corn canopy across 125 on-farm evaluation trials conducted during 4 yr.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect‡</td>
<td>CSNT</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.37***</td>
<td>0.14</td>
</tr>
<tr>
<td>CNST</td>
<td>0.000326***</td>
<td>0.0000074</td>
</tr>
<tr>
<td>Random effect</td>
<td>Location</td>
<td>(1.21)</td>
</tr>
<tr>
<td>Fixed effect</td>
<td>RGR</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–8.36***</td>
<td>0.77</td>
</tr>
<tr>
<td>RGR</td>
<td>7.62***</td>
<td>0.69</td>
</tr>
<tr>
<td>Random effect</td>
<td>Location</td>
<td>(1.01)</td>
</tr>
<tr>
<td>Fixed effect</td>
<td>GR</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>–3.10***</td>
<td>0.41</td>
</tr>
<tr>
<td>GR</td>
<td>0.031***</td>
<td>0.0038</td>
</tr>
<tr>
<td>Random effect</td>
<td>Location</td>
<td>(1.29)</td>
</tr>
</tbody>
</table>

‡ Fixed parameters shown are for the field-level models, which were estimated from intercepts that varied and slopes that were constant across 125 trials.

A low need for additional N. A vertical dashed line crossing the 0.5 line separates profitable and unprofitable areas, which would be considered as N deficient and sufficient, respectively. The vertical line crosses the x axis at about 1000 mg NO₃–N kg⁻¹, indicating that the currently used optimal category (700–2000 mg NO₃–N kg⁻¹) for the CSNT is adequate to identify the corn N status. The value of 1000 mg NO₃–N kg⁻¹ is a reference or midpoint for the estimated optimal range. The lower and upper limits for the optimal range in Fig. 6 will depend on how users consider the uncertainty in predicting a profitable YR expressed as probability values.

The model in Fig. 6A correctly predicted 75% of observations for the two binary categories of economic YR. Because the specificity percentage was larger than the sensitivity percentage, the model was slightly better in predicting unprofitable sampling areas than profitable ones. This observation is reasonable because a relatively large variation in YR was observed in the deficient range, <250 mg NO₃–N kg⁻¹ (Fig. 3E). From its maximum value of 1.0, Cohen’s kappa chance-corrected index was 0.50, indicating a satisfactory predictability of CSNT across all the trials (Fig. 6A).

The relationship between the predicted probability of a profitable YR and RGR values across all the trials was described by a steep-sloping S-shaped curve (Fig. 6B). The curve shows that the probabilities rapidly increased with higher RGR values. The model in Fig. 6B correctly predicted 78% of the observations for the two binary categories of economic YR. This prediction was slightly better than that for the stalk NO₃ values shown in Fig. 6A and with a slightly higher Cohen’s kappa index value. Similarly to the CSNT, RGR predicted slightly better unprofitable than profitable YR values.

Unlike the probability curve for the CSNT (Fig. 6A), the steep S-shaped curve for RGR makes it easier to identify a single critical value (Fig. 6B). The critical value separating N-deficient and -sufficient categories for RGR was 1.10. This critical value is in
the same range as shown in Fig. 4 when RGR values were related to YR in each year and across all 4 yr. The similarities in critical values suggest that both analyses produced similar outcomes. Unlike the analyses in Fig. 3 and 4, multilevel analysis allowed adjustments for imagery spectral properties in each trial and to quantify the uncertainty by calculating probability distributions (Fig. 6). This uncertainty, however, could be important when making management decisions based on the results of the late-season evaluations.

An S-shaped curve described how the probability of a profitable YR changed with an increase in GR values for the low N rate across 125 trials (Fig. 6C). The critical value separating N-deficient and -sufficient categories was 95. Regression predictability parameters for GR were slightly better than those for the CSNT (Fig. 6A) and slightly lower than those for RGR (Fig. 6B).

A unique feature shown in Fig. 6B when using RGR is that the observed values of an unprofitable YR tended to be located in the left lower corner of the graph and the observed profitable YR values tended to be located on the top, to the right of the critical RGR value. The ideal classification (correctly classifying 100% of samples) for RGR would locate all observed profitable YR values at the upper right and all unprofitable YR values in the lower left corner. For the CSNT, this separation was less obvious and the ideal classification for the CSNT would put all unprofitable YR values to the right of the critical value and all profitable YR values to the left. For GR of the corn canopy, a relatively large number of unprofitable areas had GR values between 100 and 200 (Fig. 6C), suggesting that this diagnostic incorrectly predicted some profitable YR values. The relatively good predictability for GR of the imagery should not be always expected and should be used with caution because the imagery in this study was enhanced and normalized. Low-cost digital aerial imagery or imagery from free sources is usually not radiometrically corrected, however, or normalized (enhanced) across fields.

The model diagnostic parameters and visual observations in Fig. 6 indicate that the three late-season diagnostic tools performed equally well, with slightly better prediction and higher Cohen’s kappa index for RGR of the corn canopy. Additional analysis showed that combining two variables (e.g., CSNT and RGR or CSNT and GR) in one model slightly decreased the AIC and residual deviance values (data not shown), but the models correctly predicted almost the same percentage of observations in economic YR values. Therefore, each of the discussed diagnostic tools can be used separately for late-season N evaluations.

Testing additional field-level covariates such as the total N rate applied to the normal N treatments and site-specific average monthly rainfall in the multilevel regressions together with CSNT, RGR, or GR (similar to Fig. 6) showed that N rates did not have a significant effect on the probability of a profitable YR (data not shown) but cumulative spring (March–May) and growing season (March–August) rainfall had small significant positive effects. After considering the effect of the CSNT with each additional centimeter of spring rainfall, the probability of a profitable YR increased by about 2%; after considering the effects of RGR or GR with each additional centimeter of growing season rainfall, the probability of a profitable YR increased by about 1%.

Estimated critical CSNT values for two categories of trials based on the previous crop (C–S or C–C) were in the range currently considered as optimal (700–2000 mg NO$_3$–N kg$^{-1}$); however, the probability of a profitable YR for C–S trials was slightly higher, especially when CSNT values were >2000 mg NO$_3$–N kg$^{-1}$ (data not shown).

**Effect of Nitrogen Cost and Corn Price on Critical Values**

One of the methods for quantifying the uncertainty in the estimated critical values for the late-season diagnostic tools is to address the effects of potential fluctuations in corn and N fertilizer prices. Although the average corn and fertilizer prices usually correlate positively, the effects of different corn/N price ratios on critical values of the CSNT are shown in Fig. 7. We used three price ratios: the average price ratio observed during the 4 yr of the study, 30% lower, and 30% higher than the average. Multilevel models were used to predict the probability of a profitable YR when the marginal cost of additional N expressed as the amount of grain was 0.22, 0.31, and 0.41 Mg per 56 kg of N. The highest number represents relatively expensive N and a relatively low price for corn; the lowest number represents relatively inexpensive N and a relatively high price for corn.

The downward sloping probability curve for the lowest (0.22 Mg per 56 kg N) marginal cost of N crossed the 0.5 probability line at about 2300 mg NO$_3$–N kg$^{-1}$, and the downward sloping probability curve for the highest marginal cost (0.41 Mg per 56 kg N) crossed the 0.5 probability line at about 200 mg NO$_3$–N kg$^{-1}$. This range in stalk NO$_3$ values is very similar to currently recommended marginal and optimal (250–700–700–2000 mg NO$_3$–N kg$^{-1}$) ranges in Iowa (Blackmer and Mallarino, 1996). The curve for the most expensive N, however, decreased more rapidly than those for the less expensive N, with the probability of a profitable YR in the deficient range (>250 mg NO$_3$–N kg$^{-1}$) being slightly higher than 0.5. Overall, the price variation did not change the critical value for the CSNT.

Discussions of the performance of the CSNT in the deficient range deserves special attention. The predicted probabilities in
a range from 0 to 250 mg NO₃–N kg⁻¹ did not exceed 0.70; however, probabilities ≤0.5 would indicate a random chance in predicting a profitable YR. Thus, it would be risky to conclude that sampling areas that tested <250 mg NO₃–N kg⁻¹ would always require additional N. These relatively low probability values can be partially explained by a relatively large variability in YR observed in the deficient range (Fig. 3) and by the fact that corn plants can remobilize all of the N from the lower portion of the stalk to the grain by the time the corn grain matures without substantially affecting corn yields or economic YR.

Figure 7 suggests that for the average marginal N cost (the solid line), the probability of a profitable YR in the deficient range compared with that in the range from 1000 to 2000 mg NO₃–N kg⁻¹ was about 50% higher (0.6 vs. 0.4), and the probability of a profitable YR in the deficient range compared with that in the range of >5000 mg NO₃–N kg⁻¹ was about three times higher (0.6 vs. 0.2). These estimates still indicate a relatively large potential payoff from an increased N supply in the deficient range and, at the same time, provide a quantitative assessment of the risk associated with interpretation of the deficient corn N status for the CSNT.

**Differences in Critical Values among Nitrogen Management Practices**

Since the CSNT was introduced in the early 1990s (Binford et al., 1990, 1992), N management practices in Iowa have gradually changed, including more frequent use of liquid swine manure and UAN. Manure is usually applied in the fall and UAN is applied in the spring (preplant) or sidedressed. To study the effect of timing and N forms on critical values for the CSNT, we grouped N management practices into four categories based on a common combination of N forms and timing of applications. The N management categories were: (i) fall-injected liquid swine manure; (ii) sidedress UAN or AA; (iii) fall-applied AA; and (iv) spring or preplant UAN.

Figure 8 shows predicted probability curves for the four N management categories studied during the 4 yr. A category of spring-applied AA was not used because of the relatively small number of trials that used this management practice (Table 1). The curve for fall manure shows a relatively large critical stalk NO₃ value of about 3000 mg NO₃–N kg⁻¹. The other N management categories of sidedress N, fall AA, and spring UAN showed critical values in the range of 200 to 700 mg NO₃–N kg⁻¹, which is the same range as currently recommended for the marginal category in Iowa. The higher critical concentration for fall manure is surprising because the average total N applied with normal manure rates for both C–C and C–S was from 20 to 40 kg N ha⁻¹ higher than the average total N rate applied to corn in the other N management categories (Table 1).

The larger estimated critical stalk NO₃ value for fall manure can be explained by a relatively larger percentage of deficient and excessive stalk samples observed in each year compared with the other management categories (data not shown). The relatively high predicted probability (about 0.65) of a profitable YR in the deficient range for fall manure (Fig. 8) could be due to potentially large N losses observed in some years and relatively early timing of manure application in the fall. Also, the higher probabilities in this range could be because of the uncertainty in manure analysis and the assumptions used when crediting manure N to corn, often resulting in N availability much lower than expected. Also, the relatively high probabilities (0.3–0.5) predicted in the excessive range (>3000 mg NO₃–N kg⁻¹) can be partially explained by higher yield potentials from manure applications or by additional N uptake at the end of the growing season. This last explanation is possible because organic N in liquid swine manure can become available to corn plants in the second half of the growing season (Balkcom et al., 2009). Even if this N becomes available later in the season, however, it often cannot offset the potential yield loss from early or midseason N deficiency. The probability curve for sidedress N seems to support the last explanation because the predicted probabilities of a profitable YR in the excessive range were much higher than those for fall AA and spring UAN (Fig. 8).

Figure 8 also shows that except for the fall manure, when stalk NO₃ values were >2000 mg NO₃–N kg⁻¹, the predicted probabilities decreased rapidly. Therefore, possible errors in stalk NO₃ values in the above-optimal range might be less important than those in the deficient or optimal ranges. The data also illustrate that manure management requires more scrutiny than other N management categories, and potential adjustments of the critical ranges are needed when making reliable interpretations for the CSNT used for fields where liquid swine manure is injected in the fall.

**CONCLUSIONS**

On-farm two-treatment evaluations can be used to verify existing categories of corn N status for the CSNT and to identify critical values and optimal ranges for the two other late-season N diagnostics based on the RGR or GR of digital aerial imagery...
of the corn canopy. Because producers are often interested in whether an additional N supply during the growing season would change the corn N status or produce a profitable YR, the proposed calibration method was based on estimating the probability of receiving a profitable YR to additional N and identifying critical values that separated profitable, probably N-deficient areas from unprofitable, probably N-sufficient areas.

All three late-season N diagnostic tools correctly predicted almost the same percentage of economic YR (profitable and unprofitable). The RGR method was slightly superior to GR and the CSNT and explained some of the observed negative YR to N. The use of RGR of the corn canopy, however, requires applying at least two N rates in each trial; the use of GR of the corn canopy requires normalizing reflectance values of uncalibrated digital aerial imagery across many corn fields.

The identified sufficient or optimal category of corn N status for the CSNT was almost the same as that currently recommended in Iowa (700–2000 mg NO3–N kg–1), even when considering a 30% deviation from the long-term average prices of corn and N fertilizer. The calibration probability curve for fall-injected liquid swine manure, however, showed a much higher critical value (about 3000 mg NO3–N kg–1) than that for fall AA, spring UAN, or sidedress N (200–700 mg NO3–N kg–1). This is probably due to the delay in N availability, relatively large N losses in some years, or the uncertainty in crediting N from fall-injected liquid swine manure.

Acknowledgments

The study was partially funded by the Iowa Soybean Association with soybean check-off dollars and by the Integrated Farm Livestock Management Project from the Soil Conservation Division of the Iowa Department of Agriculture and Land Stewardship. We are very thankful to all farmers, agronomists, and technical providers who participated in the study. This study would not have been possible without the On-Farm Network staff, who contributed numerous hours to this project.

References


Kyverga, P.M., T.M. Blackmer, R. Pearson, and T.F. Morris. 2011. Late-season digital aerial imagery and stalk nitrate testing to estimate the percentage of areas with different N status within fields. J. Soil Water Conserv. 66:373–385. doi:10.2148/jswc.66.6.373


