

## Research Article

# Estimating Uncertainty of Economically Optimum N Fertilizer Rates

**Martin Bachmaier and Markus Gandorfer**

*Center of Life Sciences, Technische Universität München, Weihenstephan, 85354 Freising, Germany*

Correspondence should be addressed to Martin Bachmaier, bachmai@wzw.tum.de

Received 13 August 2011; Revised 4 December 2011; Accepted 6 December 2011

Academic Editor: M. Ali Tabatabai

Copyright © 2012 M. Bachmaier and M. Gandorfer. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nitrogen rate trials are often performed to determine the economically optimum  $N$  application rate. For this purpose, the yield is modeled as a function of the  $N$  application. The regression analysis is commonly used to estimate the modeled functions and the economic optimum rate,  $N_{\text{opt}}$ . However, computer programs do not calculate confidence intervals for  $N_{\text{opt}}$  derived from quadratic yield response model or other commonly used models. The objective was to develop a method for computing and interpreting confidence intervals for  $N_{\text{opt}}$ . These confidence intervals can be estimated using an online program VINO.EXE. For the  $N$  rate trials on the experimental field Sieblerfeld (Bavaria), confidence intervals were computed for a range of wheat and  $N$  fertilizer prices and for selected  $N$  rates for quadratic and linear plateau models. The latter concerns the comparison with confidence intervals based upon the linear-plus-plateau yield regression model. All intervals were found to be unexpectedly wide and their ranges were affected by  $N$  rates used in the calculations and by the choice of yield response model.

## 1. Introduction

The effect of  $N$  fertilizer on the yield of agricultural crops can be studied using  $N$  response functions. Such functions are usually fitted to the data from  $N$  rate trials by regression. The available function types used for modeling purposes in the course of this discussion are, for example, quadratic (e.g., [1, 2]) or the linear-plus-plateau functions (e.g., [3]) and the Mitscherlich, which is a kind of exponential model [4, 5]. Other researchers had additionally investigated the quadratic-plus-plateau model [6, 7], the square-root model [6, 8], and even more complicated models [9]. On the basis of the assessed production functions, the economically optimum  $N$  application rate was estimated.

The economic optimum is reached when the marginal cost of the  $N$  fertilization corresponds to the marginal revenue, that is, when the returns above the  $N$  fertilizer cost (RANC) are maximized. For a given product price  $p$  (in €Mg<sup>-1</sup>) and a  $N$  fertilizer price  $r$  (in €kg<sup>-1</sup>), these returns are computed as

$$\text{RANC} = Y \cdot p - N \cdot r, \quad (1)$$

where  $Y$  is the measured yield in Mg ha<sup>-1</sup> and  $N$  is the  $N$  rate in kg ha<sup>-1</sup>. The  $N$  rate, where these returns above the  $N$  fertilizer cost are maximized, is the economically optimal  $N$  rate,  $N_{\text{opt}}$ .

The evaluation of any  $N$  rate trial is usually followed by the analysis of the residuals and the determination coefficient  $R^2$ , to justify the choice of the model applied. However, as discovered by Cerrato and Blackmer [6],  $R^2$  is not a suitable measure, as it barely depends on the chosen model. The point estimate for  $N_{\text{opt}}$ , which is derived from the fitted model does not, however, provide any information on the accuracy or reliability. Therefore, our objective is to compute and discuss confidence intervals for  $N_{\text{opt}}$ , which will be based on quadratic  $N$  response functions, with the consideration of extensive  $N$  rate trials as an example. The results can be used for optimizing decision making in nitrogen management. Furthermore, from an ecological point of view, the knowledge of these confidence intervals for  $N_{\text{opt}}$  is extremely important for identifying the corresponding  $N$  rate recommendations from the lower limit of these ranges, so that economically unnecessary  $N$  balance surpluses are avoided.

## 2. The Method for Computing the Confidence Interval

*2.1. A Quadratic Model for the Yield.* It was assumed that the expected yields,  $E(Y_i)$ , can be described by a quadratic function of the total N application rates,  $N_i$ . Therefore, the yields,  $Y_i$ ,  $i = 1, \dots, n$ , were modeled as random variables that depended on  $N_i$  in the following way:

$$Y_i = \beta_0 + \beta_1 N_i + \beta_2 N_i^2 + e_i, \quad (2)$$

$N_i$  denotes fixed levels of N application rates,  $\beta_j = 0, 1, 2$ , the fixed unknown coefficients of regression, and  $e_i$  the error variables, which are assumed to be independent and normally distributed with an expected value 0 and a common unknown variance  $\sigma^2 > 0$ . The unknown coefficients  $\beta_j$ ,  $j = 0, 1, 2$ , were estimated using the least-square estimates  $b_j$ . This requires at least three different N-levels. Otherwise, the estimates would not be unique.

*2.2. The Optimum N Application Rate and Its Point Estimate.* The economically optimal N application rate,  $N_{\text{opt}}$ , is the N rate where the expected returns above N fertilizer cost,  $E(\text{RANC}) = p \cdot (\beta_0 + \beta_1 N + \beta_2 N^2) - r \cdot N$ , are maximized. This applies to the following optimum N rate:

$$N_{\text{opt}} = \frac{r \cdot p^{-1} - \beta_1}{2\beta_2}, \quad (3)$$

$N_{\text{opt}}$  results from the N where the first derivative of the model parabola,  $E(Y) = \beta_0 + \beta_1 N + \beta_2 N^2$ , equals  $r \cdot p^{-1}$ , which is the ratio between N fertilizer and the product price.

By using the least-square estimates  $b_j$  for estimating the coefficients  $\beta_j$ , the point estimate for  $N_{\text{opt}}$  was immediately reached:

$$\hat{N}_{\text{opt}} = \frac{r \cdot p^{-1} - b_1}{2b_2}. \quad (4)$$

Note that this estimator is biased because it is not a linear combination of the unbiased least-squares estimators  $b_1$  and  $b_2$ . Fundamentally, it is a ratio. Each estimator  $b_j$  in this linear model is unbiased and thus also the estimated yield, as it is a linear combination of these coefficient estimators. The estimated N optimum The estimated N optimum in (4), however, is not a linear combination of these  $b_j$ , but a ratio in essence, so that unbiasedness no longer holds. To see this, assume for simplicity that  $r = 0$  and that  $E(b_1) = \beta_1 = 1$  and  $E(b_2) = \beta_2 = -0.5$ , so that the ratio of the expected values in (4) is equal to 1. Assume further that the variance of  $b_1$  is negligible, so that  $b_1 \approx 1$ , and consider the following two  $b_2$ -values:  $b_2 = -0.1$  and  $b_2 = -0.9$ , which are symmetric when  $\beta_2 = -0.5$ . In the former case, the ratio in (4) is around 5, in the latter case around 0.56. The mean of both values is 2.78. If  $b_2$  is additionally assumed to be always negative, it is felt that the expected ratio in (4) is greater than  $N_{\text{opt}}$ , which results, according to (3), in the value 1.

*2.3. Deriving an Exact Confidence Interval for  $N_{\text{opt}}$ .* The uncertainty of a point estimate, can be seen from a confidence interval for  $N_{\text{opt}}$ . According to Lehmann's [10] duality with

confidence intervals and tests, a confidence interval or, more generally, a confidence set for  $N_{\text{opt}}$  consists of all hypothetical values  $N_0$  whose simple null hypothesis,

$$H_0 : N_{\text{opt}} = N_0, \quad (5)$$

cannot be rejected. By using (3), the null hypothesis in (5) can be reshaped to

$$H_0 : \beta_1 + 2N_0\beta_2 = r \cdot p^{-1}. \quad (6)$$

This is a linear hypothesis, so it can be tested in the framework of general linear models using the usual  $F$ -test [11], which is a likelihood ratio test. Therefore, the corresponding confidence set is called a likelihood interval, provided that it is an interval. This usually applies but it does not need to. As mentioned before, it is formed by the set of all  $N_0$  for which  $H_0$  in (5) or (6) cannot be rejected. For those interested in the area of statistical inference, this test and the corresponding derivation of the confidence set, which can be produced by explicit mathematical formulae, is described fully in Bachmaier [12]. The confidence set is analogous to that of Fieller [13], who calculated it for the ratio of parameters, and to that of Koziol and Zielinski [14], who computed it for the maximum of a quadratic regression.

Under the assumptions made (quadratic model, independent and normally distributed homoscedastic errors), tests of linear hypotheses, such as in (6), are considered exact, so exact confidence intervals are obtained, whereas confidence intervals derived according to Weisberg (2005, sec. 6.1.2) [11], Mittelhammer et al. (2000, sec. 8.8 P. 183–185) [15] and Casella and Berger [16], which are also called Wald intervals, would only provide approximate confidence intervals. They are symmetric around the point estimate, which is not even unbiased. Therefore, they “may not accurately reflect the actual, often asymmetric, uncertainty in an estimate” [17]. Such an asymmetric situation is given here as seen with rapidly increasing fertilizer application rates. The true yield function decreases more slowly, as indicated by the quadratic model, which overestimates the yield loss due to lodging [18]. Therefore, the Delta method will not be pursued but the likelihood intervals will be used to proceed exactly as mentioned above. These intervals are not symmetric around the point estimate, so neither limit depends on the other and they can adapt better to the data around either limit. The lower limit of the confidence interval is in the focus of ecological interests, as it gives the minimum fertilization that cannot be rejected as being optimal.

*2.4. Program Implementation.* The computation of the likelihood-type confidence set for  $N_{\text{opt}}$  was implemented in the Fortran program VINO.EXE. The program CIGIGRAD.EXE could also be used. It does not specialize in optimizing N applications but it is able to compute a confidence set for the  $x$ -coordinate where the modeled regression parabola has a target gradient, which must be entered by the user. Note that, if one wants to compute a confidence set for  $N_{\text{opt}}$  using CIGIGRAD.EXE with a given N fertilizer price of  $r = \text{€}0.60 \text{ kg}^{-1}$  and a crop price of  $p = \text{€}110 \text{ Mg}^{-1}$ , one must

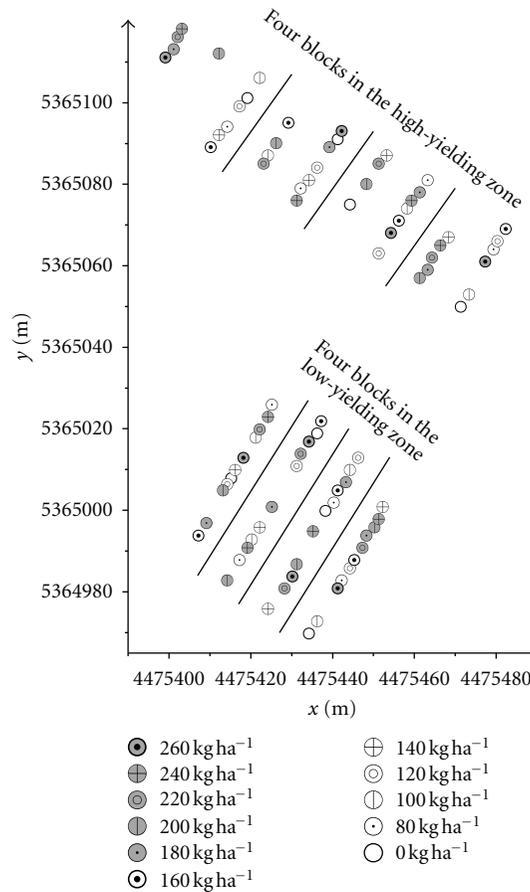


FIGURE 1: The randomized complete block design represented by the measured Gauss-Krüger coordinates  $x$  and  $y$  of the plot mid points.

not input the gradient  $r \cdot p^{-1} = 5.4545$  that results when respecting the units. Instead, only the number  $0.60 \cdot 110^{-1} = 0.0054545$ , which stands for the gradient writing  $r \cdot p^{-1} = 0.0054545 (\text{€ kg}^{-1})(\text{€ Mg}^{-1})^{-1}$  should be inserted, provided the  $N$  application rates are given in  $\text{kg ha}^{-1}$  and the yields are given in  $\text{Mg ha}^{-1}$ . A special case of the CIGIGRAD.EXE is the CIVERTX.EXE, which calculates the confidence set for the  $x$ -coordinate of the parabola vertex. This  $x$ -coordinate corresponds to the  $N_{\text{opt}}$  if the  $x$ -values are the  $N$ -levels and the  $y$ -values are the returns above  $N$  fertilizer cost,  $\text{RANC}_i$  in (1), which are to be maximized. Thus, CIVERTX.EXE can also be used for computing a confidence set for  $N_{\text{opt}}$ . All programs mentioned can be downloaded from the internet [19].

### 3. The Field Site

The test field, Sieblerfeld (5 ha), is in the Tertiary hills of Upper Bavaria, Germany, and has two very different yield zones. The soil texture in the high-yield zone is a sandy loam with an available field capacity of the rooted soil horizons of 160 mm. In the low-yield zone, the soil texture is a loamy sand with an available field capacity of the rooted soil horizons of 100 mm [20]. Within the two distinct yield zones,

$N$ -rate test areas were designated to derive site-specific  $N$ -response functions. For this,  $N$ -rate trials were carried out in the season 2001/02 on winter wheat (*Triticum aestivum* L.).

The trial design was a randomized complete block design with four blocks in each yield zone. To investigate the dependence between yield and  $N$  fertilizer rate, 11 plots that were given with different rates of  $N$  (0, 80, 100, 120, 140, 160, 180, 200, 220, 240, and 260  $\text{kg ha}^{-1}$ ) were selected randomly from each block in both zones, so that there were  $n = 44$  yield measurements in each yield zone. The yields were measured with a plot combine. The plots were  $12.5 \text{ m}^2$  with a length of 10 m and a width of 1.25 m. Figure 1, which has been adopted from Figure 1 in Bachmaier and Gandorfer [21], illustrates the randomization results at the midpoints of the plots; these were recorded with a precision of 1 m with GPS technology.

Randomized complete block designs are more beneficial compared to complete block strip trials because only the former ensures independence of the variable in question. Designs of the latter kind are widely used [22–24], but their lack of randomization within strips can result in the type of strip heteroscedasticity and correlation identified in Hurley et al. [22]. In particular, correlations within strips occur in monitored yield data [1, 24, 25] because of the threshing system's mass flow and the yield monitor's datapreprocessor.

On the other hand, complete blocking and randomization do not usually ensure complete independence, because neighboring plants at the low  $N$  rate usually rob  $N$  from the plants from the high  $N$  rate, creating the so-called border effect. Since these treatments are not much separated in space (Figure 1), the yield measured for both treatments is affected, and randomization cannot solve this problem.

## 4. Results

**4.1. Assumption Check.** In Bachmaier and Gandorfer [21], where the same trial was evaluated to test the hypothesis that there is no difference between the optimal  $N$ -rate in high- and low-yield zones (precision agriculture hypothesis) and to compute confidence intervals for the difference of the  $N$  optima the usual assumptions have been checked. The result was that the data can be assumed to be approximately normally distributed and homoscedastic, and further, a block effect in the randomized complete block design could not be detected. Therefore, the ordinary regression analysis and the mentioned methods to compute confidence sets for the  $N$  optima, can be applied, but the lack of complete independency on account of the border effect mentioned above will distort this analysis slightly.

**4.2. The Data and the Quadratic Regression Function.** Figure 2, which is also already depicted in Bachmaier and Gandorfer [21], shows the trial results and the quadratic regression curves as estimations of the yield functions for high- and low-yield zones when all  $N$  levels are considered. The adjusted coefficient of determination in high-yield sites is  $R_{adj}^2 = 0.857$ ; the low-yield sites have an  $R_{adj}^2$  of 0.726.

The narrow boxes in Figure 2 indicate the confidence intervals for  $N_{opt}$  in both yield zones, which overlap. This, however, does not mean that a statistical proof for different economic optima in both zones cannot be provided. This proof could be reached if the confidence interval for the difference of the  $N$  optima in both zones did not contain the value zero [21]. In the following subsections, only separate confidence intervals for  $N_{opt}$  in both yield zones are computed and analyzed.

**4.3. The Asymmetry of the Confidence Intervals.** Figure 2 also shows the confidence intervals for  $N_{opt}$  when the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 0.0054545 (\text{€ kg}^{-1})(\text{€ Mg}^{-1})^{-1} = 5.4545$ . This applies, for example, to  $r = \text{€}0.60 \text{ kg}^{-1}$  and  $p = \text{€}110 \text{ Mg}^{-1}$ . This ratio corresponds to the quadratic model's target slope that is to be reached by the optimum  $N$  fertilization. It is indicated by dotted lines in Figure 2.

A likelihood ratio confidence set needs not to be symmetric around the point estimate. The confidence interval  $[204 \text{ kg ha}^{-1}, 328 \text{ kg ha}^{-1}]$  for the high-yield zone is only  $36 \text{ kg ha}^{-1}$  long to the left of the point estimation, whereas to the right it is  $98 \text{ kg ha}^{-1}$ , which is nearly three times as long. This is, above all, due to the fact that yields from very high  $N$  rates do not seem to sink. For the low-yield zone, a point estimation of  $199 \text{ kg ha}^{-1}$  and a confidence interval

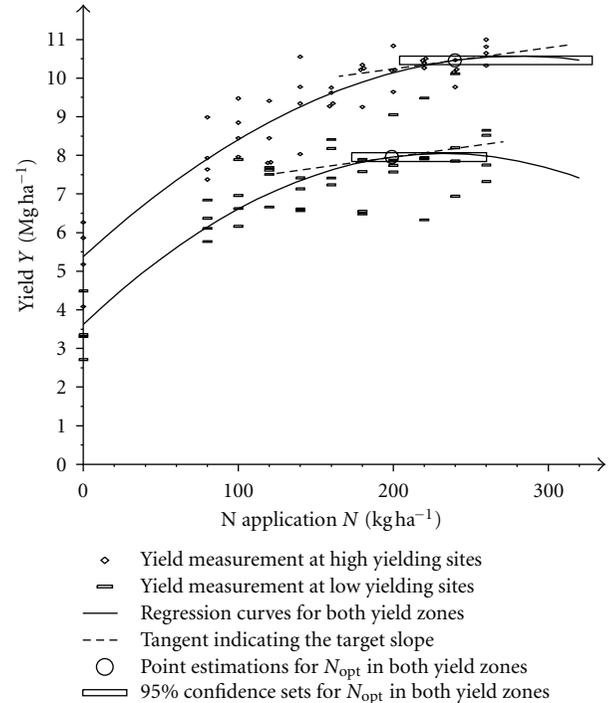


FIGURE 2: All measured values (including the zero  $N$  rate), regression curves and confidence sets for  $N_{opt}$  in both yield zones of the Sieberfeld trials, where the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

of  $[173 \text{ kg ha}^{-1}, 260 \text{ kg ha}^{-1}]$  present a similar situation with regards to the asymmetry. It seems that a concave parabola with a vertex at the far right of the point estimation is easily compatible with the measured data, whereas a concave parabola with a vertex at the far left of the point estimation is not suited for fitting the yields, since those to the right of such a vertex do not sink. Therefore, the proposed method for determining exact confidence intervals fits well to the measured data.

**4.4. The Length of the Confidence Intervals.** As can be seen in Figure 2, the 95% confidence interval for  $N_{opt}$  in the low-yielding zone  $[173 \text{ kg ha}^{-1}, 260 \text{ kg ha}^{-1}]$  and is somewhat shorter than the 95% confidence interval for  $N_{opt}$  in the high-yield zone  $[204 \text{ kg ha}^{-1}, 328 \text{ kg ha}^{-1}]$ . This can be explained as follows. In contrast to the high-yield zone, the higher  $N$  doses in the low-yield zone did not result in further increases in yield while a small yield depression could even be observed at  $260 \text{ kg ha}^{-1}$  that forced the regression function to go down. Consequently, the parabola maximum can be more clearly determined so that a long confidence interval could also be avoided. Higher  $N$  rates in the higher-yielding zone have a shortening effect on the confidence interval for  $N_{opt}$ . This will also be demonstrated in Table 2, where confidence sets are compared with regards to the different designs of the  $N$  levels. It was observed that the omission of both the maximum  $N$  application rate and zero  $N$  application rate enlarged the confidence set enormously.

TABLE 1: Effect of price ratio  $r \cdot p^{-1}$  between crop and N fertilizer on the point estimate  $\hat{N}_{opt}$  and the confidence interval (CI) and its length for different confidence levels, where the measured yields of all N levels in Figure 1 are respected.

$rp^{-1}$	$\hat{N}_{opt}$	80 % CI	length	90 % CI	length	95 % CI	length	99 % CI	length
kg ha <sup>-1</sup>									
High yield zone									
0	281	[247, 342]	95	[239, 371]	131	[234, 403]	169	[223, 505]	282
2	266	[235, 321]	86	[228, 347]	118	[223, 375]	153	[213, 467]	254
4	251	[223, 300]	77	[217, 322]	106	[212, 348]	136	[203, 429]	226
6 <sup>1</sup>	236	[211, 279]	68	[205, 298]	93	[201, 321]	120	[193, 391]	198
8	220	[199, 258]	59	[194, 275]	81	[189, 294]	105	[183, 354]	171
10	205	[186, 237]	50	[182, 251]	69	[178, 267]	88	[172, 317]	145
12	190	[174, 216]	42	[170, 228]	58	[167, 240]	74	[161, 280]	119
Low yield zone									
0	232	[208, 276]	69	[202, 297]	95	[198, 321]	123	[190, 398]	208
2	220	[198, 259]	61	[193, 277]	85	[189, 298]	109	[182, 366]	184
4	208	[188, 242]	54	[183, 258]	75	[180, 276]	96	[173, 335]	161
6 <sup>1</sup>	196	[178, 225]	47	[174, 239]	65	[171, 254]	84	[164, 303]	139
8	184	[168, 209]	41	[164, 220]	56	[161, 232]	72	[155, 273]	118
10	171	[157, 192]	35	[154, 201]	48	[151, 211]	61	[145, 243]	98
12	159	[146, 177]	30	[143, 184]	41	[140, 191]	52	[133, 214]	82

<sup>1</sup>We obtain  $rp^{-1} = 6 = 0.006 \text{ € kg}^{-1} (\text{€ Mg}^{-1})^{-1}$  if, for example,  $r = 0.66 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

TABLE 2: Comparing point estimate  $\hat{N}_{opt}$  and confidence sets at different confidence levels when considering the different designs of N levels, where the ratio between N fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

	Without max. Nrate <sup>1</sup>		All N levels <sup>2</sup>		Without zero Nrate <sup>3</sup>	
kg ha <sup>-1</sup>						
High yield zone						
$\hat{N}_{opt}$	226		240		249	
80% CI and length <sup>4</sup>	[199, 278]	79	[214, 284]	70	[213, 440]	227
90% CI and length	[193, 304]	111	[208, 305]	97	[208, 1066]	858
95% CI and length	[188, 335]	147	[204, 328]	124	$(-\infty, -657] \cup [203, \infty)$	$\infty$
99% CI and length	[180, 449]	269	[196, 402]	206	$(-\infty, 7] \cup [196, \infty)$	$\infty$
Low yield zone						
$\hat{N}_{opt}$	195		199		227	
80% CI and length	[173, 238]	65	[181, 230]	49	$(-\infty, 82] \cup [183, \infty)$	$\infty$
90% CI and length	[168, 260]	92	[177, 244]	67	$(-\infty, 130] \cup [175, \infty)$	$\infty$
95% CI and length	[164, 288]	123	[173, 260]	87	$(-\infty, 155] \cup [162, \infty)$	$\infty$
99% CI and length	[157, 404]	247	[167, 312]	145	$(-\infty, \infty)$	$\infty$

<sup>1</sup>The confidence sets are based on all measured yields at the N rates  $N_1-N_{10}$  according to Table 1.

<sup>2</sup>The confidence sets are based on all measured yields at the N rates  $N_1-N_{11}$  according to Table 2.

<sup>3</sup>The confidence sets are based on all measured yields at the N rates  $N_2-N_{11}$  according to Table 1.

<sup>4</sup>Single numbers behind intervals denote of the preceding confidence set.

When considering the enormous length of both the 95% confidence intervals, it becomes clear that even in such extensive N rate trials, the ex post estimated optimum N rates can only be roughly estimated. It is not possible, in retrospect, to limit the optimal nitrogen quantity to a level of less than 87 kg ha<sup>-1</sup> length in the low-yield zone or 124 kg ha<sup>-1</sup> in the high-yield zone. The length of the confidence intervals results from the fact that N fertilization has a very wide marginal profit area. From the economic point of view, however, the additional N quantity used eradicates

the economic advantage of the increase in yield. This leads to a very flat function around the optimum for returns above N fertilizer cost, making it compatible with many model parabolas with widely ranging vertices. The set of these vertices is the confidence interval, which is, therefore, very long.

4.5. Effects of N Fertilizer Price Increase on Confidence Intervals. Table 1 shows the dependence of point estimation and confidence intervals for  $\hat{N}_{opt}$  on the proportion between N fertilizer and crop price,  $r \cdot p^{-1}$ . The table also makes

reference to the case of  $r \cdot p^{-1} = 0$ , where the  $N$  fertilizer is gratis and the corresponding point estimates and confidence intervals for  $N_{\text{opt}}$  are point estimates and confidence intervals for the yield maximum, as it coincides with the economic optimum. Table 1 clearly shows that an increase in  $N$  fertilizer prices or, more generally, an increase in the proportion between  $N$  fertilizer and crop price,  $r \cdot p^{-1}$ , leads to a lower volume of  $N$  use.  $N_{\text{opt}}$  then would move further to the left (see Table 1). In the high-yield zone, for example, a proportion price doubling from  $r \cdot p^{-1} = 0.6$  to  $r \cdot p^{-1} = 1.2$ , which may occur when the  $N$  price doubles against an unchanged winter wheat price, reduces the point estimation from  $236 \text{ kg ha}^{-1}$  to  $190 \text{ kg ha}^{-1}$ , thus shifting the 95% confidence interval from  $[201 \text{ kg ha}^{-1}, 321 \text{ kg ha}^{-1}]$  to  $[167 \text{ kg ha}^{-1}, 240 \text{ kg ha}^{-1}]$ . The confidence interval length is reduced from  $120 \text{ kg ha}^{-1}$  to  $74 \text{ kg ha}^{-1}$ —because it is offset from the area where the true yield function seems to be relatively flat. The modeled parabolic curve for the yield  $Y$  at this  $N_{\text{opt}}$  is no longer close to the horizontal but has instead a greater positive gradient. In this region, the model parabola appears to come much closer to the real situation. However, the area where the true unknown yield function is relatively flat seems to be much wider than indicated by the fitted parabola. This was also criticized by Boyd et al. [18], who pointed out that quadratic yield functions do overestimate yield loss due to lodging, that is, true production function does not sink as quickly as a modeled parabola.

**4.6. Regression Analysis When Omitting the Zero  $N$  Application Rate.** It is justifiable to omit  $N$  levels below  $80 \text{ kg ha}^{-1}$  in order to determine the economic optimum, as it is not reached at low  $N$  rates. Because then the area to be fitted around the unknown true optimum becomes smaller, and a smaller area can be better approximated by simple functions such as quadratic ones that never model the  $N$  response precisely over a wide area.

Figure 3 shows the trial results and the quadratic regression curves for both yield zones when the results of the zero  $N$  rate are not respected. The trial results of all other ten  $N$  levels  $80, 100, 120, \dots, 260 \text{ kg ha}^{-1}$  were used for the analysis. The adjusted coefficient of determination in high-yield sites is  $R_{\text{adj}}^2 = 0.691$ ; the low-yield sites give only a low  $R_{\text{adj}}^2$  of  $0.320$ .

**4.7. The Influence of the  $N$  Rate Trial Design on the Confidence Sets.** In order to analyze the influence of the  $N$  rate trial design on the point estimate and confidence set for the economically optimum  $N$  rate, point estimates and confidence interval for  $N_{\text{opt}}$  with differing numbers of  $N$  levels were calculated. The results are shown in Table 2. Data from the low- and high-yield sites were processed separately, independent of each other.

**4.7.1. Omitting the Zero  $N$  Rate.** If zero fertilization were not taken into account in the estimate, it would lead to a higher estimation of the economically optimum  $N$  rate,  $N_{\text{opt}}$ , and to the lengthening of the confidence interval. The 95% and 99% confidence sets shown represent confidence sets that are

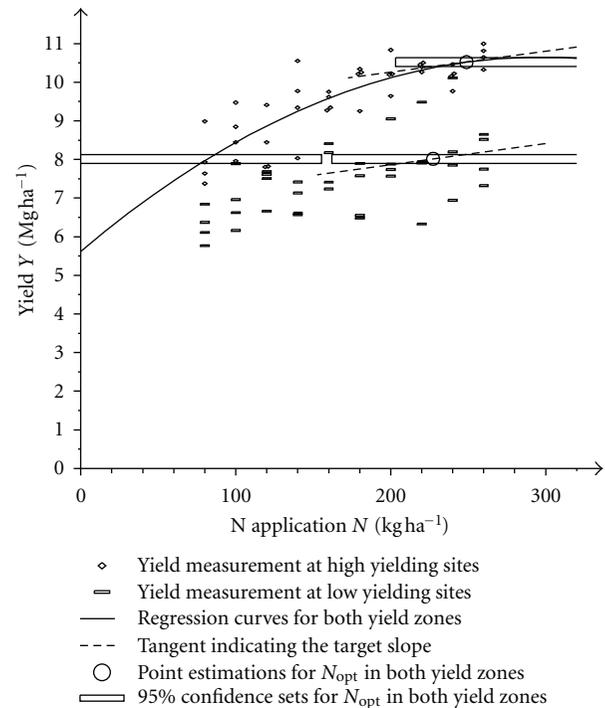


FIGURE 3: All measured values except the zero  $N$  rate, regression curves and confidence sets for  $N_{\text{opt}}$  in both yield zones of the Sieblerfeld trials, where the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

no longer intervals. They are made up of all real numbers with or even without the exception of an interval. The latter applies to the 99% confidence set in the low-yielding zone. The reason for such degenerated confidence sets is that the shape of the parabola, concave or convex, around which the data disperse, cannot be clearly recognized when the zero  $N$  rate is omitted. The data could be fitted passably by a line with a positive gradient and thus, almost equally, by very wide parabolae whose vertices lie to the far left if they are convex and to the far right if they are concave. This way, a gap-type confidence set arises, which contains  $N$ -coordinates for profit maxima (on the right of the gap) as well as profit minima (on the left of the gap).

The gap can disappear if, in addition, the error probability is so “slight” that every vertex can be considered as being “slightly” compatible with the data. In practice, all these confidence sets are, of course, worthless. Based on the research findings, feasible fertilization rules cannot be given at a high level of confidence. The gap type or the extreme length of these confidence sets, without zero fertilization, is an indication of the fact that the function for the returns above  $N$  fertilizer cost, which is to be maximized, must be relatively flat in the area of its maximum, making it very difficult to locate.

**4.7.2. Considering All  $N$  Levels.** If, however, the zero  $N$  rate is taken into account, the parabola is forced to sink as a result

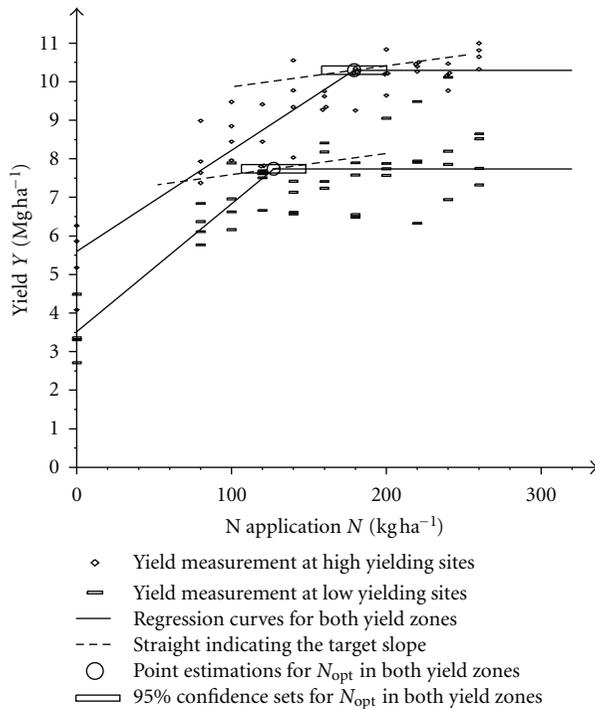


FIGURE 4: All measured values (including the zero  $N$  rate) as fitted by the linear-plus-plateau model and confidence sets for  $N_{opt}$  in both yield zones of the Sieblerfeld trials, where the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

of the low yield under no fertilization, but not just to the left of the maximum. The symmetry of the parabola means that it would sink to the right as well. In this way, the region around the maximum is less flatly modeled as it appears to be in reality. The shorter length of the confidence interval that results from this may also possibly be attributed to a weakness in the quadratic model in the area of  $N$  optimum. However, by excluding zero fertilization, the model is limited to a really small area of interest and the weakness is then of little importance. The smaller the region, the better it can be modeled by the parabola. On the one hand, the confidence interval is extended when omitting the zero fertilization, but on the other hand, it is also made more trustworthy since it is less influenced by the model assumptions.

**4.7.3. Omitting the Highest Tested  $N$  Rate.** If the highest tested  $N$  level is excluded, a minor reduction in the point estimate of the optimum  $N$  rate can be seen in comparison to when all  $N$  levels are included (Table 2). In particular, the confidence interval also extended upwards in spite of the decrease of the lower boundary and the point estimate. The highest tested  $N$  level is still in the vicinity of the  $N$  optimum. If test values are missing in this area, there would be absolute uncertainty with regard to the possible maximum on the far right. The modeling of this vertex, of a downward facing parabola to the far right, is possible if the parabola is not forced to adapt to the data in this region. As a result, the

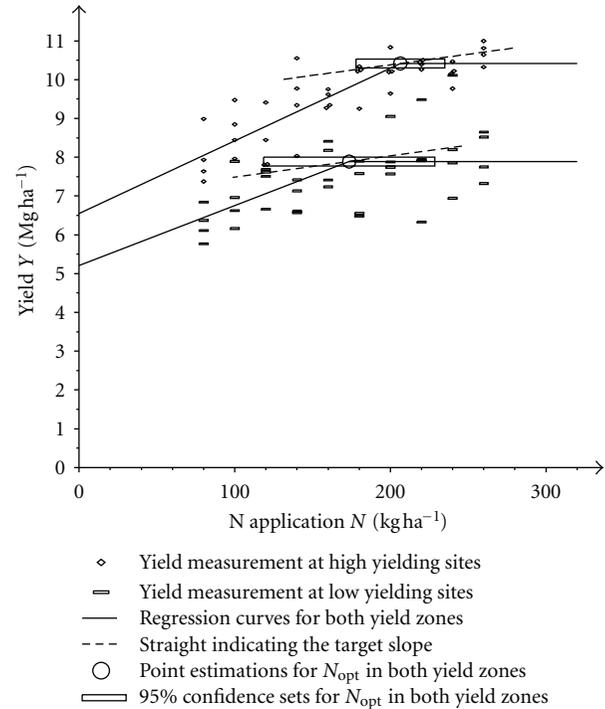


FIGURE 5: All measured values except the zero  $N$  rate as fitted by the linear-plus-plateau model and confidence sets for  $N_{opt}$  in both yield zones of the Sieblerfeld trials, where the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

confidence interval extends considerably to the right when the highest tested  $N$  level is excluded. When it comes to ensuring a sensible calculation of the confidence interval, it is essential to take high  $N$  levels into account.

**4.8. Comparison of Confidence Sets to the Linear-Plus-Plateau Model.** In the introduction, five function types used for modeling the yield response are mentioned (quadratic, linear-plus-plateau, Mitscherlich, quadratic-plus-plateau, square-root). The evaluation of  $N$  rate trials based on these models is related to the point estimates for  $N_{opt}$ . Usually, confidence intervals are only computed in the linear-plus-plateau model, where  $N_{opt}$  is one of its parameters. The confidence intervals of Hernandez and Mulla [2], however, are also based on the quadratic model, but they were not compared with those based on the linear-plus-plateau model. Yet we will see that such a comparison, which will be made hereafter, reveals large differences.

Figures 4 and 5 show the data of both yield zones as fitted by the linear-plus-plateau functions. Figure 4 takes all 11  $N$  levels into account, whereas Figure 5 excludes the zero  $N$  rate. Both figures also show the corresponding 95% confidence intervals, which were computed by the program PRISM 4.0 [26].

As far as the goodness of fit is concerned, it can be stated that both models, the quadratic and the linear-plus-plateau model, fitted equally good. Cerrato and Blackmer

TABLE 3: The comparison of the quadratic model and the linear-plus-plateau model with regards to the coefficient of determination,  $R^2$ , the point estimation,  $\hat{N}_{opt}$ , and the 95% confidence set, where the ratio between  $N$  fertilizer price  $r$  and crop price  $p$  is  $r \cdot p^{-1} = 5.4545$ , for example,  $r = 0.60 \text{ € kg}^{-1}$  and  $p = 110 \text{ € Mg}^{-1}$ .

Model	Design: all $N$ levels <sup>1</sup>				Design: all $N$ levels except the zero $N$ rate <sup>2</sup>			
	$R^2$	$\hat{N}_{opt}$	95 % CI	length	$R^2$	$\hat{N}_{opt}$	95 % CI	length
			kg ha <sup>-1</sup>				kg ha <sup>-1</sup>	
High yield zone								
Quadratic	0.863	240	[204, 328]	124	0.707	249	$(-\infty, -657) \cup [203, \infty)$	$\infty$
Lin. + plat.	0.849	179	[158, 200]	42	0.708	206	[178, 235)	057
Low yield zone								
Quadratic	0.739	199	[173, 260]	87	0.355	227	$(-\infty, 155] \cup [162, \infty)$	$\infty$
Lin. + plat.	0.715	127	[106, 148]	42	0.314	174	[118, 229)	110

<sup>1</sup> The confidence sets are based on all measured yields at the  $N$  rates  $N_1$ – $N_{11}$  according to Table 1.

<sup>2</sup> The confidence sets are based on all measured yields at the  $N$  rates  $N_2$ – $N_{11}$  according to Table 1.

[6] also made the same conclusion, but this does not validate the finding completely because, despite this similarity, Cerrato and Blackmer [6] and many other authors found marked discrepancies with respect to the  $N_{opt}$  [27–30]. Such discrepancies also apply to the confidence intervals for  $N_{opt}$  and their length. The linear-plus-plateau model provides smaller confidence sets, as can be seen in Table 3. These confidence sets do not, contrary to the parabola-based confidence sets, depend on prices unless the price ratio  $rp^{-1}$  is greater than the gradient of the increasing straight line [31]; but this is not the case. When considering all  $N$  levels, as in Figure 4, the increasing straight line's slope is  $0.026 \text{ (Mg ha}^{-1}\text{)(kg ha}^{-1}\text{)}^{-1} = 26$  for the high-yield zone and  $0.033 \text{ (Mg ha}^{-1}\text{)(kg ha}^{-1}\text{)}^{-1} = 33$  for the low-yield zone. Figure 5 depicts the design without the zero  $N$  rate, where the slope of the increasing straight line of the linear-plus-plateau model is 19 for the high- and 15 for the low-yield zone. All these slopes are greater than the realistic price ratios  $rp^{-1}$ , which at present do not exceed the value 12, as seen in Table 1. Thus, the economic optimum,  $N_{opt}$ , equals the transition of the increasing straight line to the horizontal, which is a parameter of the linear-plus-plateau model. Since PRISM calculates asymptotic confidence intervals for all model parameters [26] (2003, chap. E), they were automatically obtained for  $N_{opt}$ . They are displayed in Table 3.

The shorter length of the confidence intervals can be attributed to the fact that the transition from the increasing straight line to the horizontal is clearly easier to identify than the position where a parabola has a rather flat positive gradient. It is not surprising that the inferences from the two models, the linear-plus-plateau and the quadratic one, which are based on such very different biological assumptions, are incompatible. Nevertheless, it may be surprising that the confidence intervals for  $N_{opt}$  differed so much in both models. When considering only the design with all  $N$  levels, these confidence intervals did not even overlap. In the high-yield area, the linear-plus-plateau model resulted in a 95% confidence interval of [158 kg ha<sup>-1</sup>, 200 kg ha<sup>-1</sup>], whereas the quadratic model yielded a confidence interval of [204 kg ha<sup>-1</sup>, 328 kg ha<sup>-1</sup>]. The same pattern was seen in the low-yield area. If the linear-plus-plateau Model is

assumed to be true, then the  $N_{opt}$  estimate lies between 106 and 148 kg ha<sup>-1</sup>, whereas the quadratic model considers fertilization between 173 and 260 kg ha<sup>-1</sup> as optimal.

## 5. Discussion

In spite of the symmetrical model, the proposed confidence interval for  $N_{opt}$  is asymmetric around the point estimate. This has the advantage of better adaptation to the real situation, but it can also deviate from the interval form, so that it contains profit maxima as well as profit minima. The latter is the exact opposite of what we want. These types of degenerated confidence intervals only occur when extreme  $N$  rates, such as zero  $N$  or very high  $N$  rates are missing. When only  $N$  rates that lie in the area of the possible  $N$  optimum are tested, the yield data do not clearly indicate that the model parabola is concave. Confidence sets are only useful if they are small, but short confidence intervals can only be obtained if the empirical data clearly indicate a concave model and thus a small area where the true profit maximum could be. Then gap-type confidence sets, which also contain profit minima, would no longer occur, and the disadvantage that the proposed kind of confidence set does not differ between profit maxima and minima will not be an issue.

Therefore, it seems urgently necessary to calculate confidence intervals based only on designs that include the zero  $N$  rate and very high  $N$ , so that a concave shape of the  $N$ - $Y$  scatterplot is clearly recognizable. By using such datasets, these extreme  $N$  levels will have, due to their leverage effect, a strong influence on the regression parabola and thus also the point estimation and the confidence interval for  $N_{opt}$ . The point estimation and confidence interval are then mainly determined by the choice of the modeling of the yield function between the extreme  $N$  levels, which lie far away from the economic optimum. The estimation of  $N_{opt}$  now only seems to be a “by-product” of the fitting of an unknown yield function to a very wide range through a simple model. As seen in Table 3, the confidence sets based on different models are so different that they do not even overlap when all  $N$  levels are considered. The true model, if it exists, would lead to a good estimate of the by-product,  $N_{opt}$ , but

none of the known simple models could describe the reality accurately over a wide area.

The quadratic model is characterized by the extreme symmetry around the maximum yield, although there is nothing symmetrical in the nature, especially in the field conditions and in many ecological systems. That is why many studies in the past have shown that the quadratic model greatly overestimated  $N_{opt}$ , and also the present study shows that the quadratic model leads to much higher  $N_{opt}$  estimates than the linear-plus-plateau model. The constant slope or gradient of the latter is also not natural, but the quadratic curve is especially troublesome because  $N$  rates in the deficient range impact yield predicted at the above optimal range (by forcing yields to artificially decrease) and vice versa, even if the yield decrease from excessive  $N$  rates is not evident. If wheat lodging is observed from excessive  $N$  rates, then the quadratic model might be appropriate.

Marked discrepancies with respect to the underlying model had already been pointed out by many authors [6, 27–29]. These discrepancies equally affect the corresponding confidence intervals. Computing and comparing them with regard to different models is, therefore, an important field of research. The resulting discrepancies and the enormous length of the confidence intervals heighten further the awareness of how inaccurately  $N_{opt}$  can be estimated.

## 6. Conclusions

The usual modeling and estimation of yield functions and their  $N$  optima inspire the impression of scientific exactness. The derived point estimate of  $N_{opt}$  together with a not all too low  $R^2$  are misleading in creating overconfidence in the point estimate of  $N_{opt}$ . In this article, however, confidence intervals were derived to show that this confidence in the  $N_{opt}$  estimation and  $R^2$ , at least for these experimental data, is not advisable. The confidence intervals were extremely long and strongly dependent on  $N$  level design and modeling. Thus, the economic  $N$  optimum can only be very inaccurately estimated from yield regression models based on  $N$  rate trials. This is also attributed to the fact that the area of  $N$  fertilizer applications that are very close to the economic optimum is very wide. The economic profit would not reveal much of a difference in an  $N$  application range of 150 kg ha<sup>-1</sup> to 250 kg ha<sup>-1</sup>. Each of these values lies in approximation to the optimum. There is a very wide range in which the exact optimum could be found. Despite the fact that a particularly comprehensive  $N$  rate trial was performed, something which can only rarely be implemented in practice, such a search by means of yield data is still very difficult due to local heterogeneity and the weakness of all models.

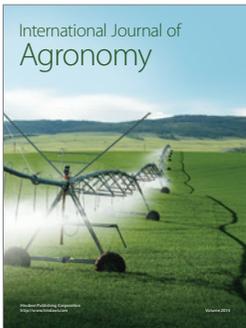
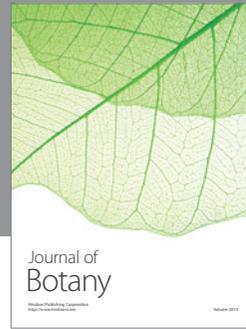
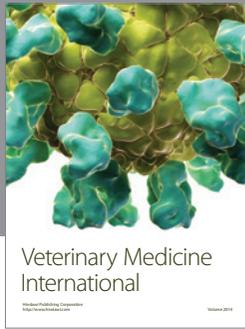
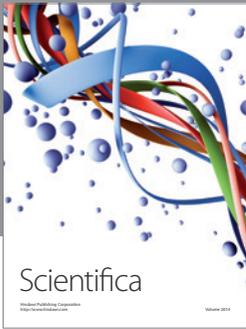
By means of the presented methods and further  $N$  rate trials, further studies should be done to see if the findings here do recur under other site conditions and experimental years. Should this be the case, then future approximations of the economic optima should be given in addition to their exact point estimations. Because these ranges are very wide, the challenge is, from an ecological point of view, to identify the corresponding  $N$  rate recommendations from the lower

limit of these ranges, so that economically unnecessary  $N$  balance surpluses are avoided.

## References

- [1] D. Lambert, J. Lowenberg-Deboer, and R. Bongiovanni, "Spatial Re-gression, an alternative statistical analysis for landscape scale on-farm trials: case study of variable rate nitrogen application in Argentina," in *Proceedings of the 6th International Conference on Precision Agriculture*, P. C. Robert, Ed., ASA/CSSA/SSSA, 2002.
- [2] J. A. Hernandez and D. J. Mulla, "Estimating uncertainty of economically optimum fertilizer rates," *Agronomy Journal*, vol. 100, no. 5, pp. 1221–1229, 2008.
- [3] P. Wagner, "Produktionsfunktionen und precision farming (Response functions in the context of Precision Farming). Zukunftsorientierte betriebswirtschaft und informationstechnologien in der agrarwirtschaft," *Gießener Schriften zur Agrar- und Ernährungswirtschaft*, vol. 29, pp. 39–66, 1999 (German).
- [4] R. M. Lark and H. C. Wheeler, "A method to investigate within-field variation of the response of combinable crops to an input," *Agronomy Journal*, vol. 95, no. 5, pp. 1093–1104, 2003.
- [5] National Academy of Sciences and National Research Council, *Statistical Methods of Research in Economic and Agronomic Aspects of Fertilizer Response and Use. Committee on Economics of Fertilizer use of the Agricultural Board*, NAS-NRC Publication 918, NAS-NRC, Washington, DC, USA, 1961.
- [6] M. E. Cerrato and A. M. Blackmer, "Comparison of models for describing corn yield response to nitrogen fertilizer," *Agronomy Journal*, vol. 82, pp. 138–143, 1990.
- [7] D. G. Bullock and D. S. Bullock, "Quadratic and quadratic-plus-plateau models for predicting optimal nitrogen rate of corn: a comparison," *Agronomy Journal*, vol. 86, no. 1, pp. 191–195, 1994.
- [8] M. Le Bail, M. H. Jeuffroy, C. Bouchard, and A. Barbottin, "Is it possible to forecast the grain quality and yield of different varieties of winter wheat from Minolta SPAD meter measurements?" *European Journal of Agronomy*, vol. 23, no. 4, pp. 379–391, 2005.
- [9] J. D. Colwell, *Estimating Fertilizer Requirements: A Quantitative Approach*, CAB International, Wallingford, UK, 1994.
- [10] E. L. Lehmann, *Testing Statistical Hypothesis*, John Wiley & Sons, New York, NY, USA, 1986.
- [11] S. Weisberg, *Applied Linear Regression*, John Wiley & Sons, New York, NY, USA, 2005.
- [12] M. Bachmaier, "A confidence set for that x-coordinate where a quadratic regression model has a given gradient," *Statistical Papers*, vol. 50, no. 3, pp. 649–660, 2009.
- [13] E. C. Fieller, "Some problems in interval estimation," *Journal of the Royal Statistical Society B*, vol. 16, pp. 175–185, 1954.
- [14] D. Koziol and W. Zielinski, "Comparison of confidence intervals for maximum of a quadratic regression function," *Listy Biometryczne (Biometrical letters)*, vol. 40, pp. 57–64, 2003.
- [15] R. C. Mittelhammer, G. G. Judge, and D. J. Miller, *Econometric Foundations*, Cambridge University Press, Cambridge, UK, 2000.
- [16] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press, Belmont, Calif, USA, 2002.
- [17] R. D. Cook and S. Weisberg, "Confidence curves in nonlinear regression," *Journal of the American Statistical Association*, vol. 85, pp. 544–551, 1990.

- [18] D. A. Boyd, L. T. K. Yuen, and P. Needham, "Nitrogen requirement of cereals," *Journal of Agricultural Science*, vol. 87, pp. 149–162, 1976.
- [19] M. Bachmaier, "Fortran Programs: 1. Confidence set for the economically optimum nitrogen fertilization in the quadratic model: VINO.EXE. 2. Confidence set for that x-coordinate where the quadratic regression model has a given gradient (incl. special case: confidence set for the x-coordinate of the parabola's vertex): CI-GIGRAD.EXE and CIVERTX.EXE," 2011, <http://www.tec.wzw.tum.de/index.php?id=46&L=1>.
- [20] J. Liebler, *Feldspektroskopische Messungen zur Ermittlung des Stickstoffstatus von Winterweizen und Mais auf heterogenen Schlägen*, Herbert Utz, Munich, Germany, 2003.
- [21] M. Bachmaier and M. Gandorfer, "A conceptual framework for judging the precision agriculture hypothesis with regard to site-specific nitrogen application," *Precision Agriculture*, vol. 10, no. 2, pp. 95–110, 2009.
- [22] T. M. Hurley, G. L. Malzer, and B. Kilian, "Estimating site-specific nitrogen crop response functions: a conceptual framework and geostatistical model," *Agronomy Journal*, vol. 96, no. 5, pp. 1331–1343, 2004.
- [23] T. M. Hurley, K. Oishi, and G. L. Malzer, "Estimating the potential value of variable rate nitrogen applications: a comparison of spatial econometric and geostatistical models," *Journal of Agricultural and Resource Economics*, vol. 30, no. 2, pp. 231–249, 2005.
- [24] L. Anselin, R. Bongiovanni, and J. Lowenberg-DeBoer, "A spatial econometric approach to the economics of site-specific nitrogen management in corn production," *American Journal of Agricultural Economics*, vol. 86, no. 3, pp. 675–687, 2004.
- [25] D. S. Bullock, J. Lowenberg-DeBoer, and S. M. Swinton, "Adding value to spatially managed inputs by understanding site-specific yield response," *Agricultural Economics*, vol. 27, no. 3, pp. 233–245, 2002.
- [26] H. J. Motulsky and A. Christopoulos, *PRISM, Version 4.0. Fitting Models to Biological Data using Linear and Nonlinear Regression. A practical Guide to Curve Fitting*, chapter 6, Graph-Pad Software, San Diego, Calif, USA, 2003.
- [27] T. P. Abraham and V. Y. Rao, "An investigation on functional models for fertilizer response studies," *Journal of the Indian Society of Agricultural Statistics*, vol. 18, pp. 45–61, 1965.
- [28] R. L. Anderson and L. A. Nelson, "A family of models involving intersecting straight lines and concomitant experimental designs useful in evaluating response to fertilizer nutrients," *Biometrics*, vol. 31, no. 2, pp. 303–318, 1975.
- [29] H. J. Barreto and R. L. Westermann, "YIELDFIT: a computer program for determining economic fertilization rates," *The Journal of Agronomic Education*, vol. 16, pp. 11–14, 1987.
- [30] L. A. Nelson, R. D. Voss, and J. T. Pesek, "Agronomic and statistical evaluation of fertilizer response," in *Fertilizer Technology and Use*, O. P. Engelstad, Ed., pp. 53–90, ASA, Madison, Wis, USA, 3rd edition, 1985.
- [31] D. L. Waugh, R. B. Cate, and L. A. Nelson, "Discontinuous models for rapid correlations, interpretation and utilization of soil analysis and fertilizer response data," in *Proceedings of the Technical Bulletin 7th International Soil Fertility Evaluation and Improvement Program*, North Carolina State University, 1973.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

