European Journal of Agronomy 170 (2025) 127744

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Evaluating boundary line fitting approaches for detecting yield-limiting factors and critical soil nutrient concentrations

C. Miti [a,b,*], A.E. Milne [b], K.E. Giller [c], R.M. Lark [a]

[a] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK
[b] Net Zero and Resilient Farming, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK
[c] Plant Production Systems, Wageningen University, P.O Box 430, 6700AK Wageningen, The Netherlands

ARTICLE INFO

Dataset link: https://CRAN.R-project.org/package=BLA, https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/11015

Keywords:
Boundary lines
Most-limiting factor
Critical nutrient concentration

ABSTRACT

Closing the crop yield gap is critical to meeting rising global food demand driven by population growth. The boundary line (BL) methodology is widely used to assess yield gaps and identify its causes. However, the lack of a standard BL fitting method can lead to inconsistencies in outputs and recommendations. This study compared four BL fitting methods, binning, BOLIDES, quantile regression (QR), and the censored bivariate normal model (cbvn), in determining the most-limiting factor and critical values (x_{crit}) across three datasets from England (Dataset 1), East Africa (Dataset 2), and a nutrient omission-trial from Ethiopia (Dataset 3). The most-limiting factor was identified using the Law of the Minimum and experimentally via omission-trials. Agreement among BL fitting methods and between BL methodology and omission-trials was tested using Cohen/Fleiss κ-statistic. The consistency of x_{crit} from BL fitting methods was assessed using the 95% confidence interval (CI) of cbvn and compared to RB209 guidelines (Dataset 1 only). Additionally, stakeholder preferences/opinions on BL fitting methods were gathered via workshops in Nairobi and Harare. Results showed BL fitting methods generally identified the most-limiting factor consistently ($\kappa > 0.4$), but inconsistencies were observed for binning and QR methods. Experimentally-determined most-limiting factors were inconsistent with BL outputs ($\kappa < 0.2$). While most x_{crit} estimates fell within the cbvn CI, deviations occurred, especially in Dataset 2. BL fitting methods often underestimated x_{crit} compared to RB209 guidelines. Stakeholder exercise showed no evidence ($p = 0.56$) against the null hypothesis of uniform ranking of BL fitting methods. The study highlights that while BL fitting methods show general consistency, discrepancies with experimentally determined results exist. Despite consistent results, cbvn is recommended for critical nutrient estimation due to its uncertainty quantification, supporting probabilistic insights for agronomic decisions.

1. Introduction

Global increase in population coupled with the adverse effects of climate change and the reduction in agricultural land due to degradation threatens future global food security (Giller et al., 2021; Kopittke et al., 2019). To counter this, agricultural production must be increased. This can be achieved by agricultural intensification through closing crop yield gaps which can avoid the need for expansion of the land area under agriculture (Godfray et al., 2010; Foley et al., 2011). The process to close the yield gap should start with its quantification and the determination of its possible causes. Management options can be identified and policies adopted that favour the identified management options (FAO and DWFI, 2015). Various methods for bench-marking yield to determine the yield gaps have been described (FAO and DWFI, 2015). One method, which has been widely used in many agronomic studies, is the boundary line methodology (Sadras, 2020). This methodology allows for bench-marking yield especially, in cases where data are collected from non-experimental settings e.g. field surveys.

The boundary line methodology was initially proposed by Webb (1972) as a tool to model the most efficient response, y, to factor, x, when these are measured in situations where other causes of variations in y occur (Sadras, 2020). The boundary line model, therefore represents the maximum response of y, for any value of x. Webb's (1972) method involves, (i) creating a scatter plot of y against x, (ii) visually selecting the points at the upper edges assuming that they are the most efficient response and (iii) fitting a model to these selected points to represent the relationship of x and y. Over the years, this simple methodology has undergone various developments that include binning (Casanova et al., 1999), boundary line determination technique (BOLIDES) (Schnug et al., 1995), quantile regression

* Corresponding author at: School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK.
E-mail address: chawezi.miti@nottingham.ac.uk (C. Miti).

https://doi.org/10.1016/j.eja.2025.127744
Received 12 January 2025; Received in revised form 16 May 2025; Accepted 16 June 2025
Available online 16 July 2025

methods, Makowski quantile regression (Makowski et al., 2007), the Bayesian segmented quantile regression (Andrade et al., 2023) and the censored bivariate normal model (cbvn) (Milne et al., 2006b). These methods can be classified into two groups, the heuristic methods (e.g. BOLIDES, Binning and quantile regression methods) which involve making subjective decisions (e.g. the quantile to regard as the boundary in quantile regression method) and the statistical methods which use strict statistical principles to fit boundary lines making them more objective and robust as they provide measures of uncertainty for the determined parameters (i.e. cbvn, Bayesian segmented quantile regression and Markowski quantile regression). For more information on the different boundary line fitting methods refer to Miti et al. (2024b).

The boundary line methodology is useful to identify appropriate agronomic practices to increase agricultural productivity. It has been applied in many studies to identify the most-limiting factors to production (Casanova et al., 1999; Fermont et al., 2009; Shatar and McBratney, 2004; Wang et al., 2015). Such information helps agronomists and agricultural land managers to prioritize which factors to address in order to improve productivity. The boundary line methodology has also been used for determination of critical nutrient concentrations (x_{crit}) to avoid limiting crop production (Andrade et al., 2023; Evanylo et al., 1987; Lark et al., 2020; Smith et al., 2024). This has economic and environmental benefits as application of excess fertilizer to soil which exceeds x_{crit} adds unnecessary financial costs of production and may also contaminate the environment (Schut and Giller, 2020). Reliable determination of the boundary line model parameters and their measures of uncertainty is therefore important if it is to be effectively used for yield gap analysis.

Despite the usefulness of the boundary line methodology in agronomy, there is still no standard agreed method for fitting a boundary line model to a dataset (Hajjarpoor et al., 2018; Smith et al., 2024; Miti et al., 2024b). The use of different boundary line fitting methods may lead to differences in the boundary line parameters and, therefore, different post-hoc interpretations and inconsistencies in agronomic recommendations. The use of statistical methods may be advantageous as they provide a more objective way of fitting the boundary line, making the procedure of boundary line fitting more consistent and also provide measures of uncertainty. However, they may be more complex to use which may make them difficult to implement for researchers with a limited statistical background (Harris and Smith, 2009). Statistical methods as opposed to heuristic methods may also require larger data sets to implement especially those that rely on the maximum likelihood approach. This means a lot of data is needed to establish the boundary with confidence. We are not aware of any studies that compare the results and interpretations for boundary line models fitted by different methods yet this type of comparative analysis is needed to give a better understanding of which methods should be recommended for future agronomic studies. This gap highlights the need for such comparative analysis, which our study addresses as a novel contribution.

The aim of this study is to compare the output of a boundary line analysis using different boundary line fitting methods for the purposes of (i) identification of the most-limiting factor, (ii) determination of critical soil nutrient concentration, and (iii) how these impact on agronomic interpretation and recommendations using dataset from three different studies. We also present results of a stakeholder consultation with agronomists and other researchers' opinions on the use of different boundary line methods as tools for yield gap analysis that were elicited during two hands-on workshops.

2. Methodology

2.1. Datasets used

Three datasets consisting of crop yields and potentially limiting factors were used for this study. Two of these datasets have previously been used in yield gap analysis studies using the boundary line methodology (Fermont et al., 2009; Lark et al., 2020).

The first dataset, Dataset 1, was assembled by AgSpace Agriculture Ltd and includes measures of wheat yield as well as soil variables pH, potassium (K), phosphorus (P), and magnesium (Mg), all of which were taken in various management units across England. For more details see Lark et al. (2020). AgSpace Agriculture Ltd uses pre-identified management zones inside each field to guide soil sampling for its clients. Each management unit served as the foundation for the sampling zone and was defined by skilled soil scientists utilizing a free survey. In each sampling zone, 24 soil cores were taken to a depth of 15 cm, and then pooled to create a bulk sample. A sub-sample was taken from the bulk sample for laboratory analyses for P, Mg, K and pH. The Olsen's method was used to extract P while K and Mg were extracted using 1M ammonium nitrate. Soil pH was measured in 1:2.5 soil to water suspension with a pH meter. The result was treated as the estimate of the sampling zone. The mean wheat yield was measured for each zone for the year 2015 to 2017. The dataset used in this study is based on measurements done in 2016. In our study, we applied the boundary line methodology to wheat yield and the variables P, K, Mg and pH.

The second dataset, Dataset 2, was compiled by Fermont et al. (2009) from farm surveys and agronomic on-farm and research station trials in a study on closing the cassava yield gap for smallholder farms in East Africa. Data on main production constraints, socio-economic settings, farm management, and cassava crop management was collected in on-farm surveys. In each field at a farm, composite soil samples (depth of 0 – 20 cm) were taken, oven-dried, sieved through a 2-mm sieve, and analysed for pH, available P, exchangeable K, Ca, Mg, total N, soil organic carbon (SOC) and soil texture. Daily precipitation data were recorded using rainfall gauges at all sites. Research technicians scored overall weed management as well as disease incidence (yes/no) and severity. The boundary line methodology was applied to this dataset by Fermont et al. (2009) to identify the most-limiting factor in each field. In our study, we applied the boundary line methodology to cassava yield and the soil pH, available P, exchangeable K, Ca, Mg, total *N* and SOC.

The third dataset, Dataset 3, was compiled by CIMMYT for wheat nutrient omission trials conducted in two zones (West Showa and Jimma) in Ethiopia in 2015 and 2016 (Craufurd, 2017). The trials comprised six nutrient management treatments, namely Control (zero fertilizer), PK (N omitted), NK (P omitted), NP (K omitted), NPK, and NPK+Ca+Mg+Zn+B laid out in a randomized complete design replicated across individual farmers' fields. Each treatment plot measured 8 meters by 8 m, with maize planted at a spacing of 75 cm (inter-row spacing) and 25 cm (intra-row spacing). The nutrients N, P and K were applied at rates of 120, 40 and 40 kg/ha respectively using urea (46%), triple super phosphate (P_2O_5) and muriate of potash(K_2O) as sources. Nitrogen was applied in three splits, the first as a basal application, the second as a topdressing 21 days after emergence, and the third was applied as a topdressing approximately 42 days after emergence. All other nutrients (P and K) were applied as basal at the time of planting. Apart from fertilizer application, all plots received the same management practices. The plots were weeded manually twice during the cropping season. First at 2 weeks after planting and the second at 4 weeks after planting. Key soil properties SOC, total N, available P, exchangeable K, exchangeable Ca, exchangeable Mg, pH, soil texture, exchangeable acidity (H + Al) and Micro-nutrients (Zn, Fe, Cu, Mn and B) were measured in each field on soil samples to depth of 20 cm before the trial. SOC was determined using the modified Walkley and Black method while the micro-Kjeldahl digestion was used to determine total N. The Mehlich-3 extraction procedure preceding inductively coupled plasma optical emission spectroscopy was used to determine available P, K, Ca, Mg and micro-nutrients (Zn, Fe, Cu, Mn and B). Exchangeable acidity (H + Al) was determined by extracting soil with 1N potassium chloride and titration of the supernatant with 0.5M sodium hydroxide. At the end of the trial, biomass and grain yields were collected from each plot. We used the data from the control plot to fit the boundary lines to determine the most-limiting factor in the absence of fertilizer application. A total of 148 data points were used. In this study, boundary line models were fitted to SOC, N, P, K, Ca, Mg and pH data.

2.2. Fitting boundary line to data

2.2.1. Initial exploratory analysis

An exploratory analysis was conducted on all the factors used to fit the boundary line models in the datasets. Since the censored bivariate normal model works on the assumption that data are from a bivariate normal distribution with a censoring boundary frontier, exploratory analysis was conducted using histograms and summary statistics to check for the assumption of the variable being from a normal distribution possibly showing censoring of the yield variable. The octile skewness, a robust measure of skewness, was used (Brys et al., 2008). Variables with an octile skewness between −0.2 to 0.2 were assumed to be from a normal distribution. Otherwise they were log-transformed to meet the assumption. Because boundary line analysis is sensitive to outliers, we identified and removed bivariate outliers—data points that deviate markedly from the central distribution in two dimensions. To detect these, we used the bagplot, a bivariate extension of the univariate boxplot based on halfspace depth (Tukey, 1975). The central 50% of the data, referred to as the 'bag', was computed following the procedure of Rousseeuw et al. (1999), and the outer fence was defined by expanding this bag by a factor of 3. Observations falling outside the outer fence were classified as bivariate outliers and excluded from further analysis. Prior evidence for a limiting yield boundary in the response to different factors was obtained using the peel cluster approach described by Miti et al. (2024c).

2.2.2. Proposed models forms

Piece-wise broken-stick boundary line models were fitted to scatter plots for response, y as a function of factor x for the variables in the three datasets, as have been used in previous studies (Andrade et al., 2023; Lark et al., 2020). This enabled the determination of critical values (x_{crit}) in the independent variable (i.e. critical nutrient concentration) from the inflection points of the models. Two model forms, linear-plateau and trapezium boundary line models were used depending on the shape of the upper bound of a data cloud. The linear-plateau is a two-piece model that consists of a linear component that represents the functional change in response, y, with incremental changes in the independent factor, x, and a plateau component that represents the section of no change in y with change in x. It is expressed as

$$y = \min(\beta_0,\ \beta_1 + \beta_2 x), \tag{1}$$

where β_0 is the plateau response, β_1 is the y-intercept and β_2 is the slope. The trapezium model is a three piece model consisting of two linear components (with positive and negative slopes respectively) separated by a plateau component. It can be expressed as

$$y = \min(\beta_0,\ \beta_1 + \beta_2 x,\ \beta_3 + \beta_4 x), \tag{2}$$

where β_0 is the plateau response, β_1 and β_3 are the y-intercepts for the two linear components, and β_2 and β_4 are the slopes with positive and negative values respectively.

2.2.3. Fitting methods

Four commonly-used boundary line fitting methods were used to fit boundary lines to the data. These were the binning methodology (Casanova et al., 1999), BOLIDES algorithm (Schnug et al., 1995), quantile regression (Baudron et al., 2019) and the cbvn (Milne et al., 2006a).

For the binning methodology, a data cloud of the response variable, y, against a factor of interest, x, was divided into 10 sections in the x-axis as done by Casanova et al. (1999). In each section, a boundary point was selected as the point that corresponds to a given percentile of y. In this study we used the 95th and 99th percentile (τ). These are the most commonly used τ values in the literature for fitting boundary lines to data, although we note that they are somewhat arbitrary. The boundary line models were then fitted to the selected boundary points by the method of least squares.

The BOLIDES algorithm operates in several steps. First, the data relating factor x and response y were arranged in ascending order with respect to x. Three benchmark x values were defined as x_{min} for the minimum value of factor x, x_{max} for the maximum value of factor x and x_{ymax} for the value of factor x that corresponds to the largest response in the dataset. The boundary points were then selected as follows. The first boundary point was selected as the point at x_{min} with the largest y. The next point was the data point at the next x value that had the largest response greater than the previous boundary point. Subsequent boundary points were selected in this way until x_{ymax} was reached. This process was then repeated starting from x_{max}, this time moving in the opposite direction. Agronomically plausible boundary line models were then fitted to the selected boundary points by the least squares method.

The quantile regression method was implemented by fitting a boundary line model of the response y as a function of x to the data conditional of a quantile, τ, which can range from $0 - 1$ (Cade et al., 1999). Parameters of the boundary model were determined by minimizing the weighted sum difference between the modelled response and actual response (Koenker and D'Orey, 1987). In this procedure, data points that lay above the modelled response are given more weight (τ) than those that fall below it ($1 - \tau$). Two quantile values, $\tau = 0.95$ and $\tau = 0.99$, commonly used in literature to fit boundary lines were considered in this study.

For the cbvn, the boundary line was fitted using the maximum likelihood approach following the procedure described by Lark and Milne (2016). In brief, this method assumes that the observed data is from a bivariate normal distribution of x and y, which has an upper censor (boundary) which y cannot exceed. Fitting the boundary line in this case involves finding the maximum likelihood of a censored bivariate normal distribution of the data given its distribution properties (means of x and y, $\boldsymbol{\mu}$ and variance–covariance matrix, $\mathbf{C}$), the parameters of the censor/boundary line, $\boldsymbol{\beta}$ and a parameters describing the measurement error in y (σ_{me}). The likelihood of the distribution is given as the product of the likelihoods for the individual data points. As the product of a large number of small likelihoods can easily underflow the numerical precision of the computer, the log of the likelihood values was taken and the sum of the log likelihoods was computed instead. Direct estimates of σ_{me}, a fixed parameter in the cbvn, were not available and were therefore, estimated using the σ_{me} likelihood profile (Royall, 2017; Lark and Milne, 2016). This was done by fixing σ_{me} at each of a set of values in turn, and remaining parameters were estimated conditional on σ_{me} by maximum likelihood. The maximized likelihoods for the sequence of values constitute a likelihood profile. The value of σ_{me} where the profile is maximized was selected as the estimate of σ_{me}.

To determine the parameters of the boundary line, an objective negative log-likelihood function was written in `R` software (version 4.4.0) (R Core Team, 2022) and the `optim()` function was used to find the parameters that minimized the objective function using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization methods. There is a tendency for optimization algorithms to settle at a local rather than global minima. To avoid this, several starting values were used for the optimization and the output with the lowest negative log-likelihood value was chosen. In addition, the uncertainty (standard error) of the parameters was determined using the Hessian matrix, $\mathbf{H}$ (which is part of the output of the `optim()` function), following the procedure described by Dobson and Barnett (2018). This is done by taking the inverse of $\mathbf{H}$ and then taking the square root of the corresponding terms on the main diagonal.

Evidence for the validity of the cbvn compared to a bivariate normal model was assessed by computing Akaike's information criteria (AIC) for the censored and uncensored bivariate normal distribution models fitted to the data (Lark and Milne, 2016). The AIC values for a model was calculated as

$$\Lambda = 2p + 2\ell, \tag{3}$$

where Λ is the criterion, p is the number of model parameters and ℓ is the maximized log-likelihood value. This information criterion compares the models on their likelihood, with a penalty for the greater complexity of the cbvn model. In any comparison the model with smallest AIC is selected. Although the AIC is not a formal significance test, if one selects the model with smaller AIC this minimizes the expected information loss through the selection decision (Verbeke and Molenberghs, 2000).

In general we selected the cbvn model if its value of AIC, Λ_{bl}, was smaller than that of the uncensored bivariate normal model, Λ_{bvn}. Further information on the strength of this evidence for the model was obtained by computing the Akaike weight for each, as done by Lark et al. (2020). If the value of Λ for the i^{th} model, Λ_i, in a set of m exceeds the minimum value of Λ_j over the set of models considered by Δ_i then one may compute the Akaike weight for that model as

$$w_i = \frac{\exp\{-\Delta_i/2\}}{\sum_{j=1}^{m}\exp\{-\Delta_j/2\}}. \quad (4)$$

The value w_i can be interpreted as the probability that the model is the best one for the variable in the sense that the Kullback–Leibler divergence is minimal for the i^{th}, over the set of m models considered (Burnham and Anderson, 2004).

2.2.4. Determination of the most-limiting factor

For each point in the three datasets, the most-limiting factor was identified according to Sprengel-Liebig Law of the Minimum (van der Ploeg et al., 1999) as done by Fermont et al. (2009) using the equation:

$$y = \min(f_1(x_1), f_2(x_2), \ldots, f_n(x_n)), \quad (5)$$

where y is the response variable and x_i are the potential limiting factors on y according to the boundary line function $f_i(x_i)$, where $i = 1, 2, \ldots, n$. The most-limiting factor is identified as i if $y = f_i(x_i)$. However, if a point has an x value greater than the critical value for all factors and has a yield gap, the cause of the yield gap was defined as "Unknown" because all the studied factors predict that yield should be maximum. The most-limiting factor was modelled using each of the four boundary line fitting methods.

For Dataset 3, this method was applied to interpret the boundary models fitted to the yield data from control plots only. An experimental determination of the most-limiting factor was also done considering the additional plots from the nutrient omission trial. The most-limiting factor was identified for each experimental farm in the omission trials by examining the yields for all non-control plots (PK, PN, NK, NPK). The most limiting factor was identified by finding, among the plots with all, or all but one, nutrient applied, which had the smallest yield. This is the nutrient which, on addition, has the largest effect on yield. In cases where the smallest yield was observed in the treatment that received all the nutrients (NPK), the limiting factor was classified as "Unknown". Additionally, if the difference between the treatment with the smallest yield and NPK is less than the critical difference at 95% confidence, we concluded that the NPK yield was not significantly large and hence classified the limiting factor as "Unknown". The critical difference with 95% confidence was determined as

$$\Delta_{\text{crit}} = 1.65 \times \sqrt{2\sigma^2_{\text{me}}}, \quad (6)$$

where Δ_{crit} is the critical difference at 95% confidence and σ^2_{me} is the variance of measurement error (determined using the likelihood profile described in Section 2.2.3). The variance of difference between two measurements is given as $2\sigma^2_{\text{me}}$ and so the standard error of difference is equal to its square root. The 1.65 is a one-tailed t-statistic value at 95% confidence. For example, given that the yields for the treatments PK(−N), NP(−K) NK(−P) and NPK are 2, 6, 5.21 and 6.5 respectively, and the Δ_{crit} is equal to 3. The most-limiting factor is identified as N because it has the smallest yield and its difference with NPK is larger than Δ_{crit}. However, if the yield for the treatment PK(−N) was 5, its difference with NPK treatment was less than Δ_{crit} (no significant difference). In this case, the most-limiting factor was classified as unknown.

Table 1
Joint proportions of rating by two boundary line fitting methods with j limiting factor categories.

		Method 1				
		factor 1	factor 2	..	factor j	**Total**
Method 2	factor 1	π_{11}	π_{12}	..	π_{1j}	$\pi_{1.}$
	factor 2	π_{21}	π_{22}	..	π_{2j}	$\pi_{2.}$
	:	:	:		:	:
	factor j	π_{j1}	π_{j2}	..	π_{jj}	$\pi_{j.}$
	Total	$\pi_{.1}$	$\pi_{.2}$	..	$\pi_{.j}$	1

2.3. Statistical analysis

2.3.1. Comparing the most-limiting factors predicted by different boundary line fitting methods

The Cohen's and Fleiss kappa (κ) statistics were used to check the agreement among the different boundary line fitting methods in the identification of the most-limiting factor. The κ statistic is a measure of inter-rater-reliability of classification methods (Fleiss, 1971). The Cohen's κ is used when two classifications are compared while the Fleiss κ is used when more than two classifications are compared. These statistics are particularly useful when dealing with categorical data or classification problems, where models are assigning items to categories or labels. It assesses whether the agreement between the models is better than what would be expected by chance alone. The κ statistic has been widely used in medicine and psychology studies, and it has also been applied in ecological studies (Monserud and Leemans, 1992; Tarkesh and Jetschke, 2012; Tang et al., 2009). It works on three basic assumptions that (a) the subjects being rated are independent (knowing the class of subject x_1 does not give any information about the class of subject x_i, where $i = 2, 3, \ldots, n$), (b) the categories of ratings are independent, mutually exclusive and collectively exhaustive, and (c) two raters operate independently. The formula for κ is given as:

$$\kappa = \frac{\pi_o - \pi_e}{1 - \pi_e}, \quad (7)$$

where π_o is the observed agreement (the proportion of times the models actually agree) and π_e is the expected agreement by chance (Fleiss et al., 2003). For a comparison of two boundary line fitting methods (though this can be extended to comparison of more than two methods) with j number of potential limiting factors modelled over n number of points, a contingency table of proportions of agreement is generated as shown in Table 1. Note that π is the proportion and not the count value.

As described by Fleiss et al. (2003), the value of π_o is calculated as:

$$\pi_o = \sum_{i=1}^{j}\pi_{ii}, \quad (8)$$

while π_e is calculated as:

$$\pi_e = \sum_{i=1}^{j}\pi_{i.}\pi_{.i}. \quad (9)$$

The κ coefficient values range from −1 to 1. A κ of 1 shows that there is perfect agreement while a κ equal to 0 indicates that the agreement is equivalent to chance. A κ of less than 0 indicates that agreement is worse than chance. This negative value suggests that the raters are not only disagreeing but are systematically disagreeing more often than what is expected by random chance alone. Generally, κ values greater than 0.75 are considered to represent excellent agreement beyond chance while those below 0.40 are considered poor agreement beyond chance. Values of κ that are between 0.40 and 0.75 are considered to represent fair to good agreement beyond chance (Monserud and Leemans, 1992).

The z-statistic is used to test null hypothesis (H_0) that the agreement is no better than would be obtained by chance agreement ($\kappa = 0$)

against the alternative hypothesis (H_a) that the agreement is different from chance agreement ($\kappa \neq 0$). The z-statistic is computed as:

$$z = \frac{\kappa - 0}{\text{se}(\kappa)}, \tag{10}$$

where se(κ) is the standard error of κ calculated based on the formula proposed by Fleiss et al. (1969):

$$\text{se}(\kappa) = \frac{1}{(1-\pi_e)\sqrt{n}} \sqrt{\pi_e + \pi_e^2 - \sum_{i=1}^{j} \pi_{i.}\pi_{.i}(\pi_{i.} + \pi_{.i})}. \tag{11}$$

The null hypothesis was rejected if the probability of getting the calculated z-value or larger was small. This process was done using the functions `kappa2()` and `kappam.Fleiss()` from the `irr` package version 0.84.1 (Gamer et al., 2012) to compare two and more than two models respectively.

The κ statistic was also used to compare the modelled most-limiting factor using boundary lines and that obtained experimentally using omission trials for Dataset 3. The assumption that outcome categories of the raters are collectively exhaustive for use of κ-statistic was not met as the omission trial had four possible outcomes (N, P, K and Unknown) while the boundary line models had eight possible outcomes (Ca, K, Mg, N, SOC, P, pH and Unknown). However, if the classifications from the boundary line methodology and omission trials are the same, we expect points that identified pH, Ca, Mg and SOC as the most-limiting factor using the boundary line models will correspond to the classification "Unknown" for the omission trial. We therefore converted the classifications pH, Ca, Mg and SOC for the boundary line methodology to Unknown to satisfy the assumption of raters being collectively exhaustive.

2.3.2. Comparing critical nutrient concentration obtained using various boundary line techniques

The critical nutrient concentrations for variables in datasets 1, 2 and 3 were computed using the boundary line fitting methods described above and the outcomes compared. The critical nutrient concentration was determined as the inflection point of the boundary line model where an increase in soil nutrient concentration does not result in an increase in yield. This was calculated using parameters obtained from the model in Eqs. (1) and (2) as:

$$x_{\text{crit}} = \frac{\beta_0 - \beta_1}{\beta_2}. \tag{12}$$

The estimates of x_{crit} by heuristic methods were compared to the estimate obtained by the cbvn method to check if they are consistent. This was done in `R` (version 4.4.0) by computing the uncertainty around the x_{crit} determined using cbvn and checking whether the x_{crit} obtained using heuristic methods were contained in this uncertainty range. To determine the uncertainty around x_{crit} from the cbvn, 10000 combinations of the boundary line parameters were generated using the `mvrnorm()` function from the `MASS` package version 7.3-65 (Ripley et al., 2013) with mean vector equal to the censored bivariate normal parameter estimates and the estimated covariance matrix, $\boldsymbol{\Sigma}$. The value of x_{crit} was computed from each simulated set of boundary line parameters. As the values of x_{crit} were not symmetrically distributed, a 95% highest density interval (HDI) (Barry, 2011) was determined around x_{crit} as a measure of uncertainty using the `hdi()` function from the `HDInterval` package version 0.2.4 (Meredith and Kruschke, 2020). The HDI is the interval which contains the required mass such that all points within the interval have a higher probability density than points outside the interval. The HDI is particularly useful for summarizing uncertainty of a variable with an asymmetric distribution. A 95% HDI, means that there is a 95% probability that the true value of the parameter lies within that interval. Once the 95% HDI of x_{crit} was determined, we checked whether the x_{crit} obtained using the heuristic methods were contained within this range.

For Dataset 1, the estimated x_{crit} values were further compared to the standard index concentrations for soil nutrients proposed by RB209 (AHDB, 2023) for fertilizer advice in the UK. The RB209 guidelines classifies laboratory soil analysis result ranges for P, K and Mg into soil indices for each nutrient according to cereal yield response in field trials (AHDB, 2023). For P, K and Mg, yield response is expected in index 0 and 1 with application of P, K and Mg fertilizer. In index 2, 2- and 2 for P, K and Mg respectively, yield response to addition of P, K and Mg is not expected as this is the optimum range. It is therefore, expected that the x_{crit} values determined from the boundary line methodology for each nutrient to fall in these respective indices. At index 3 and above, application of P, K and Mg is not required.

2.4. Stakeholder feedback exercises

It is important that methods for yield gap analysis are accessible to scientists who may wish to use them. For this reason workshops were organized at CGIAR centres in Kenya and Zimbabwe. Participants with various specialities and levels of experience in statistical analysis were drawn from various institutions including International Institute of Tropical Agriculture (IITA), International Centre of Insect Physiology and Ecology (ICIPE), Alliance of Bioversity International and CIAT, Kenya Agricultural and Livestock Research (KALRO), University of Nairobi, French Agricultural Research Centre for International Development (CIRAD) and International Maize and Wheat Improvement Centre (CIMMYT). These were recruited by CGIAR research leads in Harare (CIMMYT) and Nairobi (IITA) who contacted potential users of the boundary line methodology for yield gap analysis. The criteria used to identify potential participants was that they work in research in agricultural crop production. A total of 32 participants took part in the workshops, 14 in Kenya and 18 in Zimbabwe. A comprehensive yield gap analysis was demonstrated to the participants using the four boundary line fitting methods, binning, BOLIDES, quantile regression and cbvn, in what was then a prototype of the BLA library (Miti et al., 2024a). `R` scripts for each of the boundary line fitting methods were provided and analysis was demonstrated on Dataset 1. Participants were then given Dataset 2 to carry out the yield gap analysis independently using the different boundary line fitting methods.

An anonymous structured questionnaire was given to participants at the end of the session to elicit their views on the various boundary line fitting methods and yield gap analysis in general. Each questionnaire had a consent form attached for participants to indicate informed consent to participate in the study. A positive ethical option (SBREC202324010FEO - SB202223/35 PhD - Miti (Lark)) was provided by the School of Biosciences Research Ethics Committee (University of Nottingham) prior to this study taking place. The questionnaire was composed of two sections. The first section consisted of questions intended to collect information on participants' research areas, their interaction with farmers and level of statistical experience. The second part contained questions to collect information on their past use of the boundary line methodology, their preferred methods (through ranking) and their views on boundary line methodology for yield gap analysis in general. The complete structure of the questionnaire is provided in Appendix D.

The question in section two required participants to rank the boundary line fitting methods presented in the workshop according to which they found easier to use from the easiest (rank 1) to the least easy (rank 4). The dataset was tested for uniformity in ranking to see if there was influence of location, statistical experience and past use of boundary line methods. When a ranking dataset is uniform, it means that all possible rankings have the same probability of being observed. To test the uniformity of ranking in the dataset, a hypothesis test was conducted on the mean rank according to the procedure of Marden (1996). Under the null hypothesis of uniform ranking for k items, the expected value of the mean rank is given as:

$$\mu_{\text{rank}} = (k+1)/2, \tag{13}$$

Table 2
Parameters of fitted boundary lines of the potential limiting factors for Dataset 1.

Method	Factor	Model	β_1	β_2	β_0	β_3	β_4	Λ_{bl}	Λ_{bvn}	ω_{bv}	ω_{bl}
Binning ($\tau = 0.95$)	P	trapezium	7.31	1.89	12.14	19.18	−1.82	–	–	–	–
Binning ($\tau = 0.95$)	Mg	trapezium	5.52	5.31	12.13	26.48	−2.67	–	–	–	–
Binning ($\tau = 0.95$)	pH	linear-plateau	18.88	3.53	11.97	–	–	–	–	–	–
Binning ($\tau = 0.95$)	K	linear-plateau	–	–	12.24	15.60	−0.67	–	–	–	–
Binning ($\tau = 0.99$)	P	trapezium	4.74	3.38	13.34	22.16	−2.32	–	–	–	–
Binning ($\tau = 0.99$)	Mg	trapezium	−15.17	8.66	13.32	34.46	−3.95	–	–	–	–
Binning ($\tau = 0.99$)	pH	linear-plateau	24.05	5.44	13.34	–	–	–	–	–	–
Binning ($\tau = 0.99$)	K	trapezium	−1.52	3.42	13.61	23.18	−1.77	–	–	–	–
BOLIDES	P	trapezium	4.76	3.46	13.57	108.35	−21.26	–	–	–	–
BOLIDES	Mg	trapezium	−18.24	9.34	13.94	40.94	−5.14	–	–	–	–
BOLIDES	pH	linear-plateau	35.01	10.21	13.97	–	–	–	–	–	–
BOLIDES	K	trapezium	−76.01	21.16	13.95	42.86	−5.04	–	–	–	–
Q.reg ($\tau = 0.99$)	P	trapezium	7.82	2.03	13.37	27.41	−3.46	–	–	–	–
Q.reg ($\tau = 0.99$)	Mg	trapezium	−15.34	8.50	13.34	37.37	−4.47	–	–	–	–
Q.reg ($\tau = 0.99$)	pH	linear-plateau	17.59	2.17	13.36	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	K	trapezium	2.04	2.53	13.42	41.31	−4.76	–	–	–	–
Q.reg ($\tau = 0.95$)	P	trapezium	7.83	1.65	12.14	15.26	−0.85	–	–	–	–
Q.reg ($\tau = 0.95$)	Mg	trapezium	−8.37	6.07	12.14	26.11	−2.58	–	–	–	–
Q.reg ($\tau = 0.95$)	pH	linear-plateau	13.74	1.12	12.13	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	K	trapezium	−6.62	5.80	12.10	44.10	−5.23	–	–	–	–
cbvn	P	trapezium	5.75 (0.88)	2.82 (0.37)	13.43 (0.09)	82.57 (12.66)	−15.77 (2.71)	**32376.46**	32 429.55	1.0	0.0
cbvn	Mg	trapezium	−15.33 (2.61)	8.13 (0.84)	13.20 (0.19)	40.62 (1.90)	−5.15 (0.31)	**33 145.73**	33 189.39	1.0	0.0
cbvn	pH	linear-plateau	26.93 (1.88)	6.98 (0.81)	13.61 (0.10)	–	–	**41 073.00**	41 093.24	0.92	0.08
cbvn	K	trapezium	−21.88 (0.04)	8.12 (0.16)	13.35 (0.19)	46.23 (7.03)	−5.67 (1.16)	**28 464.52**	28 476.13	0.99	0.01

cbvn, censored bivariate normal model; Q.reg, quantile regression model; Λ_{bl} and Λ_{bvn}, AIC values for the censored and uncensored bivariate normal models; ω, Aikaike weight for the fitted model.
Bold values of Λ_{bl} and Λ_{bvn} indicate the smaller of the two values.

and the evidence against this null hypothesis is measured using the test statistic:

$$\frac{12n}{k(k+1)} \sum_{i=1}^{k} \{m_i - \mu_{\text{rank}}\}^2, \tag{14}$$

where m_i is the mean rank of the i^{th} item, and n is the number of rankings. Under the null hypothesis, this statistic is distributed as χ^2 with $n-1$ degrees of freedom (Lee and Yu, 2013). Note that whilst participants ranked the methods 1 (favoured) to 4 (least favoured), mean rank was evaluated on reversed ranking i.e. the method which was ranked as the easiest to use was given a weight of 4 while the second, third and fourth easiest were given weights of 3, 2 and 1 respectively. Only those responses which had complete ranking provided by the respondent were considered for this test.

3. Results

3.1. Fitted boundary lines to data sets

The summary statistics and plots for the variables in the datasets used in this study are provided in Appendix A. From the initial exploratory analysis, the variable P, K and Mg from Dataset 1, SOC, K, P, Mg, Ca and N from Dataset 2 and, P and Ca from Dataset 3 were transformed to natural logs to make the assumption of underlying normality plausible. For pH in Dataset 1, the assumption of normality was achieved with Yeo-Johnson transformation (Yeo and Johnson, 2000). The test for evidence of a limiting boundary (Miti et al., 2024c) in the datasets showed that a boundary was likely (small p-values) in Dataset 1 for the variables P, Mg, K and pH while for Datasets 2 and 3, none of the variables exhibited any evidence of a boundary (See Appendix B).

All the four boundary line fitting methods were applied for yield responses to variables in Datasets 1 and 2. For Dataset 3, the binning methodology was not applied because the selected boundary data points could not represent the boundary well with the proposed bin sizes (10 bins) for all the variables. The standard deviation of the measurement error, a fixed parameter in the cbvn, was estimated to be 0.4 t ha^{-1}, 0.6 t ha^{-1} and 0.15 t ha^{-1} for Datasets 1, 2 and 3 respectively using the likelihood profiling methods (see Appendix C). The various model parameters for Dataset 1, 2 and 3 for the various boundary line fitting methods are presented in Tables 2, 3 and 4 respectively. The standard errors of the parameters are given in the brackets for the cbvn. The AIC values for the censored (Λ_{bl}) and uncensored bivariate normal (Λ_{bvn}) model for variables in Datasets 1, 2 and 3 all indicated that the boundary model was appropriate for all the variables in the datasets as the Λ_{bl} was always less than Λ_{bvn} except for the variables N and Mg in Dataset 3. However, the Aikaike weights for the boundary line models of these two variables showed a larger probability (close to 50%) for the boundary line model. The determined critical values for variable which were transformed are converted back to the original scale (see Fig. 1).

3.2. Comparison of the modelled most-limiting factors

The boundary line parameters reported in Section 3.1 were used to model the most-limiting factors for each point in the datasets using Eq. (5) described in Section 2.2.4. Fig. 2 shows the proportions of the identified most-limiting factors modelled by the different boundary line fitting methods for the three datasets. For Dataset 1, the identified most-limiting factor proportions for K, Mg, P and pH were similar for all methods. However, the binning method at $\tau = 95$ identified K and pH as most-limiting factors more often than the other methods. This difference is also reflected by the limiting factor being unknown for a smaller proportion of the variables for the binning method ($\tau = 0.95$). For Dataset 2, the binning($\tau = 0.95$) and quantile regression($\tau = 0.95$) methods generally differed from the other methods in identification of the most-limiting factor, particularly for N, P, pH and SOC. In instances where other methods did not identify a limiting factor, the quantile regression($\tau = 0.95$) largely allocated it to N and pH. For binning ($\tau = 0.95$), the smaller proportion of P identified as most-limiting factor compared to other methods was compensated by a larger proportion of N and SOC as the identified most-limiting factor.

For Dataset 3, the proportions of the identified most-limiting factors were similar for the four boundary line fitting methods. However, there was a notable difference for the proportion of Mg identified as the limiting factor for the BOLIDES method. While other methods identified Mg as the most-limiting factor in about 5% of the cases, the BOLIDES identified it in 60% of the cases. This was also reflected by a smaller

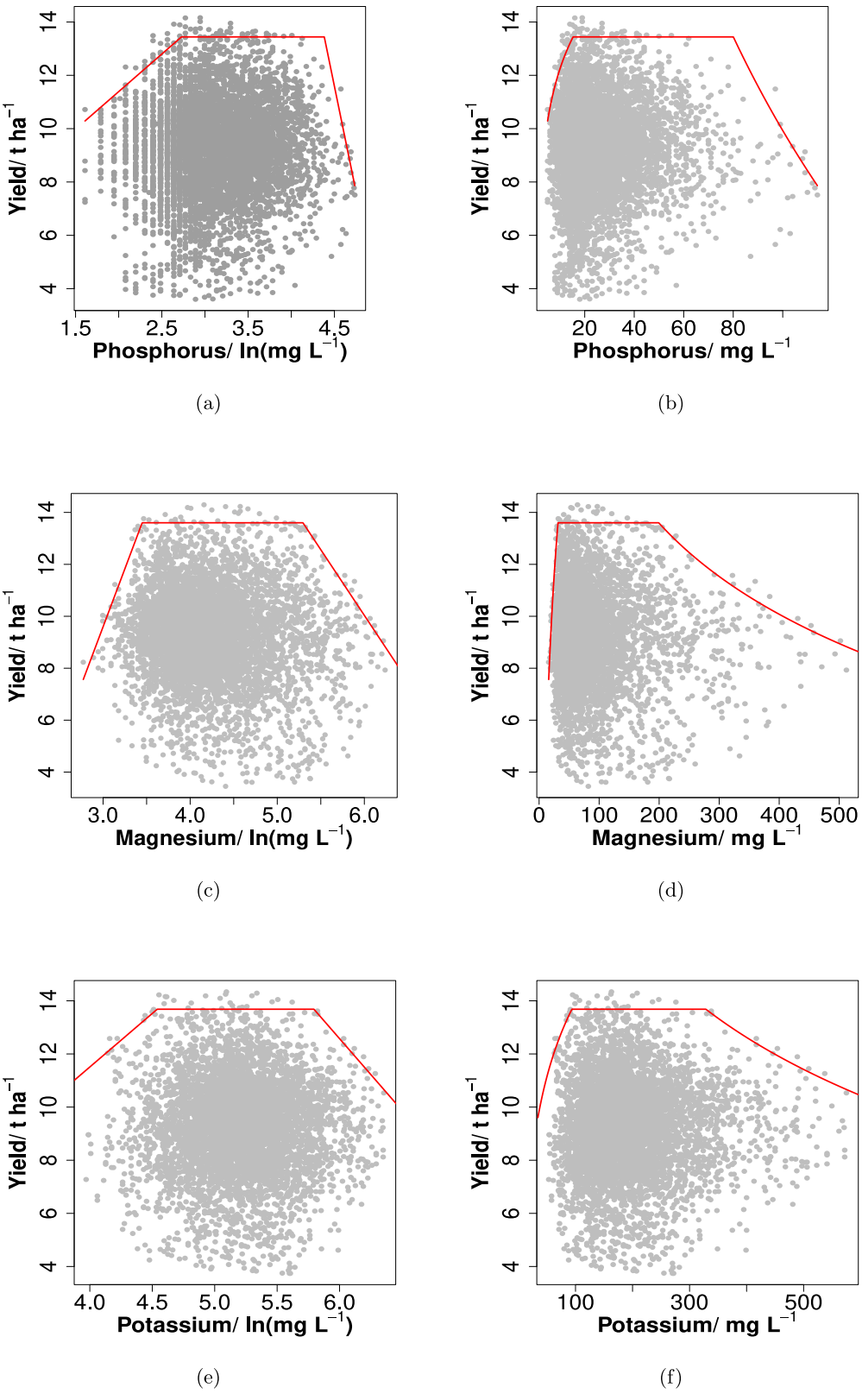

Fig. 1. Boundary lines fitted using the censored bivariate normal model for the variables P, Mg, and K from Dataset 1, displayed on the log-transformed scale (panels a, c, and e) and the original scale (panels b, d, and f) respectively.

allocation to the unknown limiting factor for BOLIDES (5%) compared to the other methods (about 60%).

Table 5 shows the κ values for comparison of the identified most-limiting factors modelled by different boundary line fitting methods with the corresponding z-statistic values and p-values for Datasets 1, 2 and 3. For Dataset 1, overall agreement (when all models were compared) was 0.52 which is fair to good agreement. This is similar to when only heuristic methods were compared (κ = 0.50). The comparison

Table 3
Parameters of fitted boundary lines of the potential limiting factors for Dataset 2.

Method	Factor	Model	β_1	β_2	β_0	β_3	β_4	Λ_{bl}	Λ_{bvn}	ω_{bl}	ω_{bvn}
Binning ($\tau = 0.95$)	SOC	Linear Plateau	21.75	41.86	22.37	–	–	–	–		
Binning ($\tau = 0.95$)	N	Linear Plateau	49.68	10.21	25.07	–	–	–	–	–	–
Binning ($\tau = 0.95$)	pH	Linear Plateau	−232.37	50.48	23.44	–	–	–	–	–	–
Binning ($\tau = 0.95$)	P	Trapezium	−1.87	33.97	23.71	35.40	−5.73	–	–	–	–
Binning ($\tau = 0.95$)	K	Linear Plateau	67.18	21.98	23.41	–	–	–	–	–	–
Binning ($\tau = 0.95$)	Mg	Linear Plateau	36.48	7.43	23.52	–	–	–	–	–	–
Binning ($\tau = 0.95$)	Ca	Linear Plateau	29.12	6.76	22.50	–	–	–	–	–	–
Binning ($\tau = 0.99$)	SOC	Linear Plateau	22.56	43.87	23.44	–	–	–	–	–	–
Binning ($\tau = 0.99$)	N	Linear Plateau	58.38	12.40	25.03	–	–	–	–	–	–
Binning ($\tau = 0.99$)	pH	Linear Plateau	−263.79	57.05	24.81	–	–	–	–	–	
Binning ($\tau = 0.99$)	P	Trapezium	0.29	27.99	25.64	45.29	−9.10	–	–	–	–
Binning ($\tau = 0.99$)	K	Linear Plateau	74.95	24.92	24.55	–	–	–	–	–	–
Binning ($\tau = 0.99$)	Mg	Linear Plateau	40.57	8.47	24.16	–	–	–	–	–	–
Binning ($\tau = 0.99$)	Ca	Linear Plateau	32.34	8.00	23.23	–	–	–	–	–	–
BOLIDES	SOC	Linear Plateau	20.50	32.10	26.17	–	–	–	–	–	–
BOLIDES	N	Linear Plateau	56.54	10.97	27.05	–	–	–	–	–	–
BOLIDES	pH	Linear Plateau	−260.39	55.44	26.66	–	–	–	–	–	–
BOLIDES	P	Trapezium	4.24	19.36	26.81	58.82	−13.89	–	–	–	–
BOLIDES	K	Linear Plateau	63.53	21.66	27.05	–	–	–	–	–	–
BOLIDES	Mg	Linear Plateau	36.24	6.91	26.81	–	–	–	–	–	–
BOLIDES	Ca	Linear Plateau	29.25	5.80	26.67	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	SOC	Linear Plateau	19.91	10.75	26.09	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	N	Linear Plateau	61.99	11.83	25.50	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	pH	Linear Plateau	−254.82	54.42	25.25	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	P	Trapezium	4.61	19.86	25.83	50.29	−10.54	–	–	–	–
Q.reg ($\tau = 0.95$)	K	Linear Plateau	73.56	25.07	25.25	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	Mg	Linear Plateau	43.64	8.38	25.25	–	–	–	–	–	–
Q.reg ($\tau = 0.95$)	Ca	Linear Plateau	30.56	6.33	25.50	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	SOC	Linear Plateau	20.10	18.51	27.33	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	N	Linear Plateau	53.17	9.05	27.33	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	pH	Linear Plateau	−254.83	54.42	26.53	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	P	Trapezium	4.46	20.98	27.33	50.29	−10.54	–	–	–	–
Q.reg ($\tau = 0.99$)	K	Linear Plateau	74.36	25.36	26.60	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	Mg	Linear Plateau	44.10	8.48	27.33	–	–	–	–	–	–
Q.reg ($\tau = 0.99$)	Ca	Linear Plateau	30.58	6.33	27.33	–	–	–	–	–	–
cbvn	SOC	Linear Plateau	20.58 (0.59)	32.23 (1.82)	26.33 (0.42)	–	–	**790.10**	799.48	0.99	0.01
cbvn	N	Linear Plateau	61.84 (2.18)	11.82 (0.59)	26.48 (0.43)	–	–	**887.53**	893.40	0.95	0.05
cbvn	pH	Linear Plateau	−255.49 (12.62)	54.53 (2.54)	26.48 (0.42)	–	–	**852.25**	858.15	0.95	0.05
cbvn	P	Trapezium	4.47 (0.67)	20.08 (0.84)	26.48 (0.44)	49.51 (12.44)	−10.24 (4.84)	**879.60**	893.61	0.99	0.01
cbvn	K	Linear Plateau	54.99 (0.15)	15.31 (0.29)	26.47 (0.44)	–	–	**934.84**	935.44	0.57	0.43
cbvn	Mg	Linear Plateau	44.09 (1.40)	8.48 (0.36)	26.48 (0.44)	–	–	**1060.17**	1065.42	0.93	0.07
cbvn	Ca	Linear Plateau	29.54 (1.42)	6.07 (0.75)	26.33 (0.43)	–	–	**1086.76**	1092.32	0.94	0.06

cbvn, censored bivariate normal model; Q.reg, quantile regression model; Λ_{bl} and Λ_{bvn}, AIC values for the censored and uncensored bivariate normal models; ω, Aikaike weight for the fitted model.
Bold values of Λ_{bl} and Λ_{bvn} indicate the smaller of the two values.

Table 4
Parameters of fitted boundary lines of the potential limiting factors for Dataset 3.

Method	Factor	Model	β_1	β_2	β_0	Λ_{bl}	Λ_{bvn}	ω_{bl}	ω_{bvn}
BOLIDES	N	linear-plateau	−45.19	396.10	9.05				
BOLIDES	pH	linear-plateau	−37.12	9.23	9.25				
BOLIDES	SOC	linear-plateau	−8.99	9.42	9.05				
BOLIDES	P	linear-plateau	6.19	2.29	9.17				
BOLIDES	Ca	linear-plateau	−5.58	8.77	9.17				
BOLIDES	Mg	linear-plateau	1.67	2.75	9.02				
BOLIDES	K	linear-plateau	3.51	7.64	9.05				
Q.reg ($\tau = 0.99$)	N	linear-plateau	0.14	47.34	9.25				
Q.reg ($\tau = 0.99$)	pH	linear-plateau	−35.02	8.81	9.25				
Q.reg ($\tau = 0.99$)	SOC	linear-plateau	−3.96	6.66	9.25				
Q.reg ($\tau = 0.99$)	P	linear-plateau	6.48	2.38	9.25				
Q.reg ($\tau = 0.99$)	Ca	linear-plateau	−5.58	8.77	9.25				
Q.reg ($\tau = 0.99$)	Mg	linear-plateau	4.84	1.49	9.25				
Q.reg ($\tau = 0.99$)	K	linear-plateau	3.51	8.29	9.25				
Q.reg ($\tau = 0.95$)	N	linear-plateau	0.15	47.33	8.66				
Q.reg ($\tau = 0.95$)	pH	linear-plateau	−31.21	7.98	8.65				
Q.reg ($\tau = 0.95$)	SOC	linear-plateau	−3.96	6.66	8.67				
Q.reg ($\tau = 0.95$)	P	linear-plateau	6.43	2.27	9.02				
Q.reg ($\tau = 0.95$)	Ca	linear-plateau	−5.53	8.72	8.65				
Q.reg ($\tau = 0.95$)	Mg	linear-plateau	1.43	3.06	8.67				
Q.reg ($\tau = 0.95$)	K	linear-plateau	3.51	8.29	8.67				
cbvn	N	linear-plateau	0.14 (0.22)	47.34 (0.08)	9.06 (0.24)	99.25	**98.76**	0.44	0.56
cbvn	pH	linear-plateau	−36.95 (0.01)	9.19 (0.04)	8.97 (0.20)	**377.62**	382.35	0.91	0.09

(*continued on next page*)

Table 4 (*continued*).

Method	Factor	Model	β_1	β_2	β_0	Λ_{bl}	Λ_{bvn}	ω_{bl}	ω_{bvn}
cbvn	SOC	linear-plateau	−8.07 (0.06)	9.13 (0.11)	8.99 (0.22)	**391.44**	394.22	0.8	0.2
cbvn	P	linear-plateau	6.33 (0.25)	2.35 (0.11)	9.06 (0.22)	**482.91**	485.55	0.79	0.21
cbvn	Ca	linear-plateau	−5.58 (3.73)	8.77 (2.52)	9.07 (0.23)	**337.68**	337.77	0.51	0.49
cbvn	Mg	linear-plateau	1.44 (1.71)	3.05 (0.95)	8.98 (0.22)	506.20	**505.37**	0.4	0.6
cbvn	K	linear-plateau	3.41 (0.39)	8.27 (1.04)	9.07 (0.26)	**374.68**	377.80	0.83	0.17

cbvn, censored bivariate normal model; Q.reg, quantile regression model; Λ_{bl} and Λ_{bvn}, AIC values for the censored and uncensored bivariate normal models; ω, Aikaike weight for the fitted model.
Bold values of Λ_{bl} and Λ_{bvn} indicate the smaller of the two values.

Table 5
The κ coefficients for most-limiting factors modelled by different boundary line fitting methods for Dataset 1, 2 and 3.

Dataset	Model comparison	raters	κ	z-statistic	p-value
1	All models	6	0.52	258.00	<0.01
1	heuristic models	5	0.50	203.00	<0.01
1	cbvn and BOLIDES	2	0.64	78.00	<0.01
1	cbvn and Binning ($\tau = 0.99$)	2	0.55	64.00	<0.01
1	cbvn and Q.Reg ($\tau = 0.99$)	2	0.77	92.60	<0.01
1	cbvn and Binning ($\tau = 0.95$)	2	0.32	51.20	<0.01
1	cbvn and Q.Reg ($\tau = 0.95$)	2	0.63	77.20	<0.01
1	Binning ($\tau = 0.99$) and Binning ($\tau = 0.95$)	2	0.39	54.70	<0.01
1	Q.Reg ($\tau = 0.95$) and Q.Reg ($\tau = 0.99$)	2	0.67	81.60	<0.01
2	All models	6	0.59	52.60	<0.01
2	heuristic models	5	0.55	41.20	<0.01
2	cbvn and BOLIDES	2	0.82	17.50	<0.01
2	cbvn and Binning($\tau = 0.99$)	2	0.82	17.50	<0.01
2	cbvn and Binning($\tau = 0.95$)	2	0.40	8.58	<0.01
2	cbvn and Q.Reg ($\tau = 0.99$)	2	0.81	17.20	<0.01
2	cbvn and Q.Reg ($\tau = 0.95$)	2	0.39	9.08	<0.01
2	Binning($\tau = 0.99$) and Binning($\tau = 0.95$)	2	0.46	10.10	<0.01
2	Q.Reg ($\tau = 0.95$) and Q.Reg ($\tau = 0.99$)	2	0.40	9.80	<0.01
3	heuristic models	3	0.37	14.10	<0.01
3	cbvn and BOLIDES	2	0.31	14.00	<0.01
3	cbvn and Q.Reg ($\tau = 0.99$)	2	0.86	18.40	<0.01
3	cbvn and Q.Reg ($\tau = 0.95$)	2	0.93	18.90	<0.01
3	Q.Reg ($\tau = 0.95$) and Q.Reg ($\tau = 0.99$)	2	0.86	18.30	<0.01
3	Omission trial and BOLIDES	2	0.01	0.31	0.75
3	Omission trial and Q.Reg($\tau = 0.99$)	2	0.00	0.17	0.87
3	Omission trial and Q.Reg($\tau = 0.99$)	2	0.00	0.00	0.99
3	Omission trial and cbvn	2	0.00	0.04	0.97

between the cbvn and any of the heuristic methods produced fair to good agreements with the exception of binning ($\tau = 0.95$) which had poor agreement. For the methods that require a subjective choice of τ, the comparison of the results using the two values had poor agreement for binning and fair to good for quantile regression. For Dataset 2, the overall agreement was fair to good when all methods were compared and similar when all the heuristic methods were compared. Good to excellent agreement was observed for comparison of the cbvn and each of the heuristic methods. The comparison between the two quantile regression methods as well as the two binning methods ($\tau = 0.95$ and $\tau = 0.99$) had fair to good agreements. For Dataset 3, only the cbvn, BOLIDES, quantile regression ($\tau = 95$) and quantile regression ($\tau = 99$), were compared. The overall agreement was fair to good agreement when all methods were compared. A poor agreement was observed when only heuristic methods were compared. The comparison between cbvn and any of the heuristic methods had excellent agreements except for the BOLIDES method which had poor agreement. The comparison of the two quantile regression methods produced excellent agreement.

Comparing the identified most-limiting factor using the boundary line methodology and the omission trials for Dataset 3, poor agreements were observed ($\kappa < 0.1$ for all comparisons). The p-values were large indicating that these κ were not significantly different from zero. The omission trial identified N as the most-limiting factor in 71.6% of the cases, P in 12.8% of the cases, K in 3.4% of the cases and 12.2% of the cases did not identify any of the studied factors as most limiting (Fig. 2(d)). On the other hand, the boundary line methods on average identified N as limiting factor in 1.4% of cases, P in 15.7% of cases, K in 3.2% of cases, Mg in 19.3% of cases, Ca in 0.8% of cases, SOC in 7.8% of cases, pH in 3.2% of cases and 48.6% of cases were unknown (Fig. 2(c)).

3.3. Comparison of critical nutrient concentration from various boundary line techniques

Table 6 shows the critical values associated with the variables in the three datasets. The upper and lower bounds of 95% confidence intervals (highest density interval) determined by the cbvn are also given. Note that the critical values for the transformed variables were back transformed to the original scale after fitting the boundary lines for easy comparison with recommended RB209 index values.

For all datasets, the critical values obtained using the quantile regression methods fell within the 95% CI of the cbvn in most instances except for Mg, K and pH in Datasets 1 and SOC in Dataset 2. However, the critical Mg values just fell at the border of the lower limit of the CI. The values obtained using the BOLIDES were consistently within the 95% CI of cbvn except for few instances mostly in Dataset 2. The critical values obtained using the binning methods mostly fell outside the 95% CI of the cbvn. Comparing the estimated critical nutrient values for Dataset 1 with the RB209 guideline indices, the critical P concentrations determined by heuristic methods all fell in the upper section of index 1 with the exception of the quantile regression ($\tau = 0.99$) which, like the cbvn, fell in the lower section of index 2. However the 95% CI of cbvn stretches further into index 2 (Fig. 3(a)). The critical Mg concentrations fell in the lower to middle section of index 1 for all methods except for the binning ($\tau = 0.95$) which fell in index 0. It is worth noting that for the variables Mg and K, there was a small number data points in the index 0. The fitted boundary line models for K using the binning and quantile regression methods at $\tau = 0.95$ (i.e initially plateau then linear model with negative slope) indicate that all values of K were above the critical concentration i.e no value of

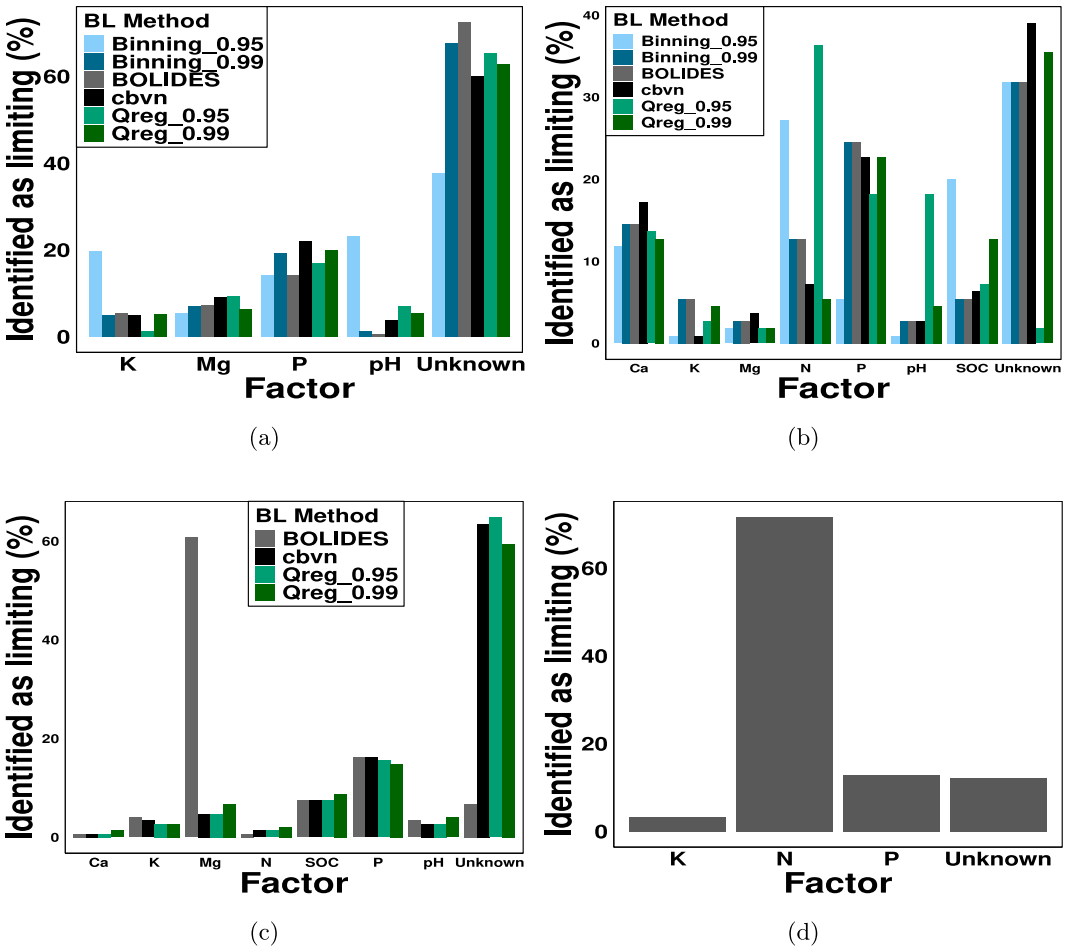

Fig. 2. Proportions of identified limiting factors modelled using four boundary line fitting methods for (a) Dataset 1, (b) Dataset 2 (c) Dataset 3, and (d) limiting factor identified using nutrient omission trials (Dataset 3).

Table 6
Critical values obtained from boundary line parameters and the upper and lower limit from the censored bivariate normal model.

Data	var	Units	cbvn	Lower	Upper	BOLIDES	Bin95	Bin99	QR95	QR99
1	P	mg L^{-1}	15.29	12.65	17.80	12.76	12.88	12.74	13.63	15.39
1	Mg	mg L^{-1}	33.42	30.38	37.47	31.36	**3.47**	**26.84**	**29.34**	**29.20**
1	pH	–	6.06	5.93	6.23	**5.84**	5.99	5.97	6.00	**6.59**
1	K	mg L^{-1}	76.61	66.19	90.94	70.21	–	83.43	**25.22**	89.84
2	SOC	g kg^{-1}	1.20	1.13	1.27	1.19	**1.02**	**1.02**	**1.77**	**1.48**
2	N	%	0.05	0.05	0.06	**0.07**	**0.09**	**0.07**	0.05	0.06
2	pH	–	5.17	5.14	5.20	5.18	**5.07**	**5.06**	5.15	5.17
2	P	mg kg^{-1}	2.99	2.81	3.19	**3.21**	**2.12**	**2.47**	2.91	2.98
2	K	cmol kg^{-1}	0.16	0.13	0.19	0.19	0.14	0.13	0.15	0.15
2	Mg	cmol kg^{-1}	0.13	0.10	0.15	**0.26**	**0.18**	0.14	0.11	0.14
2	Ca	cmol kg^{-1}	0.59	0.42	0.89	0.64	**0.38**	**0.32**	0.45	0.60
3	N	%	0.19	0.17	0.21	**0.14**	–	–	0.19	0.18
3	pH	–	5.00	4.87	5.17	5.03	–	–	5.03	5.00
3	SOC	%	1.87	1.75	2.01	1.92	–	–	1.98	1.90
3	P	mg kg^{-1}	3.20	2.47	4.10	3.67	–	–	3.21	3.14
3	Ca	cmol kg^{-1}	5.32	4.73	6.38	5.38	–	–	5.43	5.08
3	Mg	cmol kg^{-1}	2.47	2.04	3.56	2.68	–	–	2.95	2.37
3	K	cmol kg^{-1}	0.68	0.58	0.82	0.73	–	–	0.69	0.62

All values are on the original measurement scale.
Values in **bold** are outside the 95% confidence interval of the censored bivariate normal model.
RB209 P Index: **0**, 0–9 mg L^{-1}; **1**, 10–15 mg L^{-1}; **2**, 16–25 mg L^{-1}; **3**, 26–45 mg L^{-1}; **4**, 46-70 mg L^{-1}
RB209 Mg Index: **0**, 0–25 mg L^{-1}; **1**, 26–50 mg L^{-1}; **2**,51–100 mg L^{-1}; **3**, 101–175 mg L^{-1}; **4**, 176–250 mg L^{-1}
RB209 K Index: **0**, 0–60 mg L^{-1}; **1**, 61–120 mg L^{-1}; **2**–, 121–180 mg L^{-1}; **2**+ , 181–240 mg L^{-1}; **3**, 241–600 mg L^{-1}

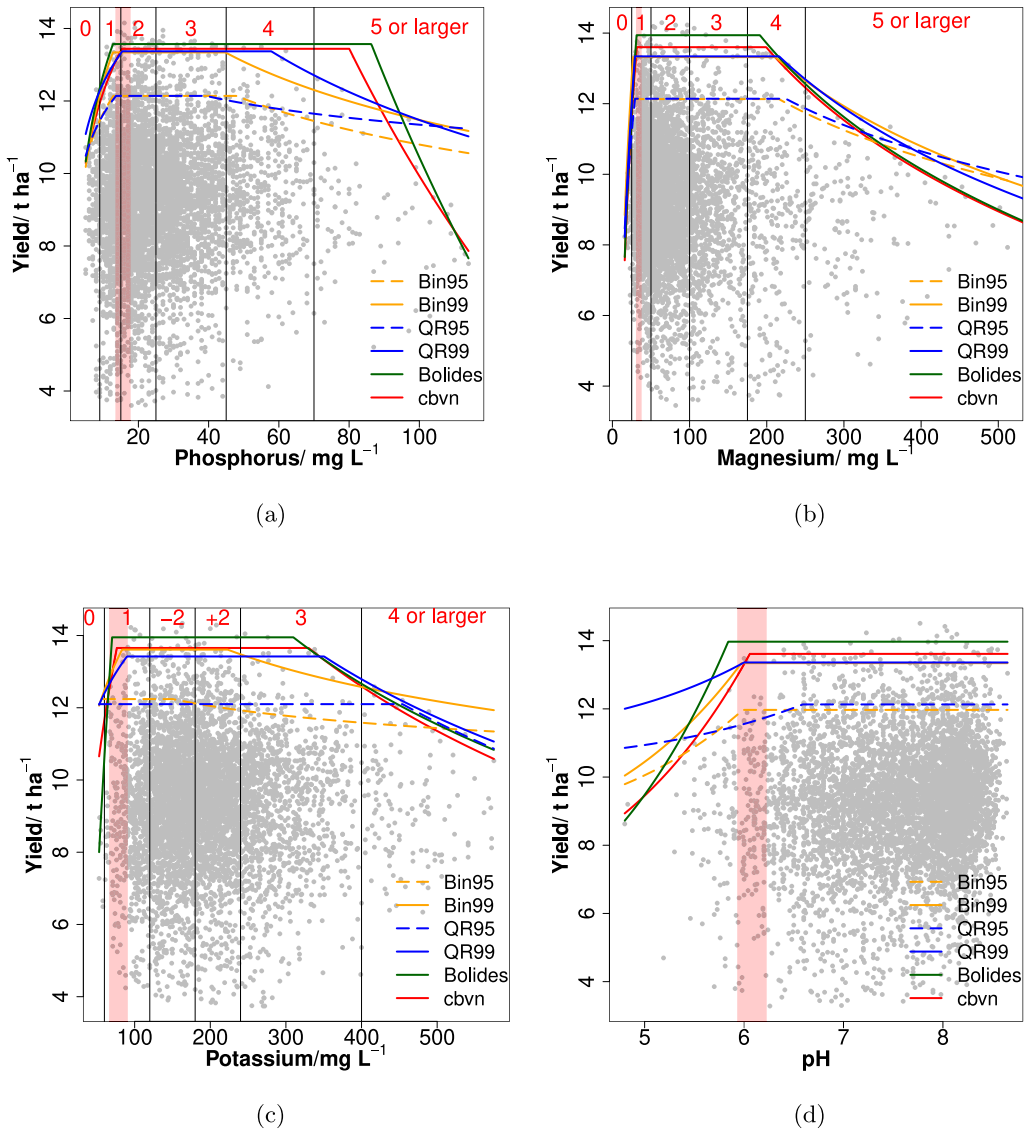

Fig. 3. Boundary line models for (a) soil P, (b) soil Mg, (c) soil K and (d) Soil pH fitted using the various methods. The reddish area is the 95% CI of x_{crit} from the cbvn model. The black vertical lines mark the boundaries used for the RB209 index categories: 0, 1, 2, 3, 4, and 5.

measured K could have limited yield by deficiency. According to the RB209 guidelines, the optimum pH for mineral soils of England is 6.5. This falls just outside the upper bound of the 95% CI of cbvn.

3.4. Stakeholder preferences

Fig. 4 shows the mean ranks of stakeholders' preferences for the different boundary line fitting methods from the study. The mean ranks for binning, BOLIDES, cbvn and quantile regression for all participants were 3.17, 2.45, 2.31 and 2.07 respectively. The test statistic (χ^2 = 2.71, df = 3) for all the participants showed that there was no evidence (p = 0.56) against the null hypothesis of uniform ranking. For adequacy of output from the different boundary line fitting methods for yield gap analysis (Fig. 4(b)), 4 participants stated all methods, 7 binning, 1 BOLIDES and 20 cbvn. Twenty three participants stated that information on uncertainty was an important output for interpretation while 9 did not give an opinion. For the adequacy of the boundary line methodology in general for yield gap analysis (Fig. 4(c)), 23 participants stated that it is adequate, 7 stated it is not adequate and 2 did not give an opinion.

The results from the elicitation exercise in the Nairobi and Harare workshops did not indicate any special preference of a particular method. The ranking of the boundary line fitting methods was uniform for the whole dataset and hence we did not further split the data by locations and statistical experience for further analysis. However, some useful information for the reasons of their ranking was obtained. Some participants stated that the input parameters for the cbvn may be unavailable and difficult to estimate (especially measurement error) as reason for ranking it lower. However, the information on uncertainty of parameters was recognized as an important output for boundary lines using the cbvn. While more participants stated that the boundary line methodology was sufficient for yield gap analysis as a whole, some participants stated that it is inadequate as it does not incorporate the interaction effects on biological response, and in the cases of the cbvn method, cannot be applied to categorical variables including social economic factors.

4. Discussion

The comparison of boundary line fitting methods for yield gap analysis was undertaken to check the consistency among the different methods to model the most-limiting factor and critical nutrient concentrations. The results for the three datasets show the importance of

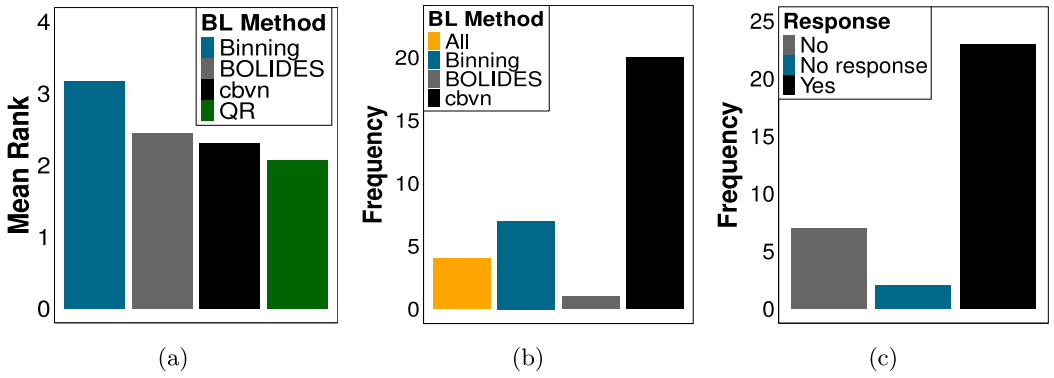

Fig. 4. (a) Ranking of the boundary line fitting methods according to user-friendliness for all participants, (b) Adequacy of output for interpretation (c) Adequacy of boundary line methodology as a whole for yield gap analysis.

considering whether a boundary model is justified, rather than automatically applying it. This is not generally done, and heuristic methods offer no basis for assessing whether the model is justified on the basis of its fit. One could, after all, draw lines through the upper fringes of any bivariate plot, and this is what the peel method aims to address at the exploratory stage, and what the AIC assessments of the cbvn model does more powerfully. Unlike heuristic methods, the statistical cbvn approach fits both bounded and unbounded models to the data, enabling model comparison. By evaluating AIC values, this method assesses whether the additional boundary parameters improve model fit. A lower AIC for the bounded model indicates that the boundary is statistically justified, supporting its interpretation. Without this step, any interpretation of the boundary would lack statistical validation and could be misleading. In addition, one can also check for evidence of clustering in the upper edges of a data cloud as justification to carry out boundary line analysis.

Dataset 1 showed evidence for boundary structure from the peel tests (*sd* mostly smaller than the bootstrap mean, and some small *p*-values) (see Appendix B) and this was supported by the AIC assessments with decisive evidence for the cbvn model with Akaike weights all above 0.9 and mostly above 0.99 (Table 2). With Dataset 2, the evidence from the peel tests was mostly positive, but larger *p*-values(Appendix B). This is consistent with smaller size of dataset and observations on limited range for some variables. Dataset 3 mostly provides no evidence for boundary structure (*sd* larger than bootstrap mean in 5 out of 7 cases for left bound and all cases for right) and assessment of the AIC rejects boundary model for N and Mg, and is much less decisive for other variables with Akaike weights for cbvn less than 0.9 for all but pH (Table 4). This shows that the AIC assessment of the cbvn model is more powerful than the peel test, but is consistent with it. The small size of Dataset 3 makes it difficult to establish evidence of boundary existence as there are less data points in the upper edges of the data cloud. These findings indicate that results based on boundary line models from Datasets 2 and 3 should be interpreted with caution, as there is no clear evidence of boundary structures in these datasets.

Agronomically plausible models were fitted to the datasets by accessing the upper boundary data structure in relation to the possible underlying biophysical and agronomic theories as emphasized by Sadras (2020). For example in Dataset 1, a piecewise trapezium model was fitted for the relationship of Mg and wheat yield. This is because an increase in soil Mg concentration might be expected to lead to increases in yield until the critical concentration above which yields do not increase further (AHDB, 2023). In general, soil Mg concentrations larger than the critical concentration do not damage crop growth, but may hinder the uptake of other cations such as K. So, where soil Mg is excessively high (index 4 and above of the RB209 guidelines), a reduction in yield may be observed due to limitation of other nutrients (PDA, 2017). This was consistent with our fitted model as the second inflection point fell within the range of index 4.

Apart from the *piece-wise* linear models, other complex models like the quadratic, logistic (Nelder, 1961; Oliver, 1964), double-logistic (Shabani et al., 2018), inverse logistic (Helidoniotis et al., 2011), monomolecular (Draper and Smith, 1998), Gompertz (Draper and Smith, 1998), Weibull (Myers and Myers, 1990), Schmidt (Schmidt et al., 2000) and Mitscherlich (Mitscherlich, 1909) models among others may be used to fit the boundary lines. In this study, we used the *piece-wise* linear models to enable the determination of the critical values at the inflection points. In cases where other complex models are used, they may produce similar fits and the simplest model should be adopted (Bargagli Stoffi et al., 2022). However, differences in model structure/form has been identified as one of the sources of uncertainty in determination of yield gap from process models (Schils et al., 2022). This may affect the boundary line model outputs in a similar way. There is need for further studies to evaluate the sensitivity of model form on the output of the boundary line methodology.

The κ statistic was used to check how consistent boundary line fitting approaches are in modelling the most-limiting factor (Tables 5). It is important to note that the κ statistic is a measure of consistency and not correctness. In general, the approaches are consistent in the determination of the most limiting factor but some inconsistencies were observed for the comparisons between the binning methods in Dataset 1 and, the comparison of quantile regression ($\tau = 0.95$) and cbvn in Dataset 2. These inconsistencies highlight the influence of the subjective decisions in the implementation of the heuristic binning and quantile regression methods, where the boundary is defined based on an arbitrary choice of τ. This emphasizes the need for the use of more objective methods to fit boundary line models if consistency is to be achieved. There was one notable instance for Dataset 3, where soil Mg was found to be limiting in substantially more cases using the BOLIDES than the other methods (Fig. 2(c)). This was also reflected in the poor agreement ($\kappa < 0.4$) for the heuristic methods. This may have been due to the effect of the data size and the arrangement of data points at the upper edge of the scatter. Unlike all other methods, the BOLIDES fits the boundary line model to the extreme points that cover the upper part of the cloud and therefore, has a higher sensitivity to the distribution of the boundary points. This means that a few extreme points, including points that are not considered as outliers, could determine the model parameters when the BOLIDES method is applied.

In the case of the omission trial, the limiting factors identified from the experimental results were not consistent with those identified from the boundary line methodology ($\kappa < 0.4$) (Table 5). This highlights the need for further work to examine the interpretation of boundary line models for yield gap analysis. As has been noted previously by Miti et al. (2024b), the boundary line need not be interpreted exclusively in terms of a limiting factor model in terms of Sprengel-Liebig, where factors have independent effects on the crop response, and only one can be limiting. However, before considering further experimental work

it is necessary to consider the potential reasons for these observed inconsistencies and the two issues which these results raise.

First, the results for exploratory analysis of Datasets 2 and 3 did not provide evidence for boundary structures in the data. In the case of Dataset 2 the comparison of the cbvn model with a null bivariate normal model did indicate that the boundary model was appropriate for all soil variables apart from Potassium (Table 3), with the Akaike weight for the cbvn model exceeding 0.93 in all cases. In contrast, for Dataset 3 from the omission trial (Table 4), the censored bivariate normal model was favoured for soil N and Mg. For other variables, the evidence for the boundary model was limited, with Akaike weights ranging from 0.51 (soil Ca) to 0.91 (soil pH). These results show weak evidence that the data exhibit a boundary which can be confidently modelled and interpreted to identify the most-limiting factor. In particular, there was no evidence for the boundary model in the case of soil N, whereas the largest group of sites in the omission trial was that for which N application had the largest effect on yield.

The small size of Dataset 3 likely contributed to the observed result. It has been emphasized that boundary line analysis requires relatively large datasets to ensure that any bounding behaviour is exhibited, and that there are sufficient observations near the bound to allow robust estimation of its parameters (Sadras, 2020; Shatar and McBratney, 2004; Miti et al., 2024c). Miti et al. (2024c) used simulated data from different distributions to examine the power of the peel cluster test to detect bounded behaviour; they found that a minimum of 400 data points were required. In the case of Dataset 3, the exploratory analysis and analysis with the cbvn model would provide grounds for not proceeding with an interpretation based on the boundary line. In most studies which use heuristic methods to fit the boundary line (Miti et al., 2024b) such an assessment cannot be made.

Another consideration, highlighted by the results for Dataset 3, is whether the appropriate soil variables have been measured to identify a limiting effect which would respond to a particular input. For a macronutrient such as P and K it is widely assumed that measurements of available nutrient concentration in the soil can be used to determine a quantitative recommendation for P or K fertilizer respectively. In the case of soil N, laboratory measurements of either total N or mineral N do not relate simply to a measure of soil N supply which is commensurate with rates of application of fertilizer N. Nitrogen in the soil undergoes constant transformations such as mineralization, nitrification and denitrification which are driven by microbial activity, temperature, moisture and other environmental factors, all of which influence the availability of plant-accessible N (Pruthviraj et al., 2024). It is therefore unlikely that a single soil N measurement will necessarily be indicative of the presence or absence of an N limitation on crop yield. It is notable that there was some evidence, albeit not strong, for a boundary line relationship between yield and SOC in Dataset 3, with an Akaike weight of 0.8. As mineralization of SOC is the primary source of the soil N supply in the absence of artificial fertilizer, this boundary relationship might be indicative of nitrogen limitation. It is necessary to consider carefully what soil measurements are most appropriate to identify and act on yield gaps.

The boundary line fitting approaches are generally consistent in the estimation of critical values. Therefore, the use of one method is not expected to produce results that differ markedly from others. In most instances, the critical values for variables estimated using the heuristic methods all fell within 95% CI of the cbvn or just outside the CI (e.g. Mg in Dataset 1 and N using the BOLIDES in Dataset 3) (Table 6). Most of the cases where the critical values fell outside the 95% CI of cbvn were observed in Dataset 2 especially when the binning methods were used. This may have been due to the combined effect of bin size, τ value considered as boundary and small size of the dataset. From this, we recommend that the boundary line methodology is applied to sufficiently large datasets with enough coverage and sufficient distribution of all possible conditions, and especially those that fulfil the assumption of boundary existence as earlier pointed out.

Generally, results from the different boundary line fitting methods are inconsistent with the RB209 guideline indices with the exception of the cbvn and quantile regression ($\tau = 0.99$) methods for soil P (Table 6). In most instances, the boundary line fitting methods underestimated the critical values placing them in index 1 and in a few case index 0. In these indices we still expect addition of P, K and Mg to increase yield (AHDB, 2023). The inconsistencies in results obtained using the different τ values for the binning and quantile regression methods indicate the impact of the subjective decision on the outcome of the boundary line analysis when heuristic method are used. Andrade et al. (2023) similarly found that varying the τ value for quantile regression had a large effect on the determined critical soil nutrient values. This stresses the need for a more objective way to determine the τ value to consider as the boundary. Makowski et al. (2007) proposed a method to determine a specific τ for quantile regression methods using expert knowledge and previous knowledge on the distribution of measurement errors using Bayesian statistics. However, this is only possible when there is enough information available on the uncertainty of the crop yield being studied. Despite, the inconsistency of the boundary line methodology with standard RB209 indices, similar studies have found that the boundary line methodology is an effective tool for understanding yield response to soil conditions (Smith et al., 2024; Andrade et al., 2023). In these studies, the obtained critical values fell within the recommended guidelines for soils of those regions.

The results from our study indicate that there is little difference in modelled limiting factors and critical nutrient concentrations using the different boundary line fitting methods. However, the statistical cbvn method offers a more objective and consistent approach to fitting boundary line models, with added robustness through uncertainty estimates for the critical nutrient concentration and a strong test for boundary presence. Schut and Giller (2020) and Andrade et al. (2023) highlighted the importance of measures of uncertainty in soil analysis measurements for field-specific fertilizer recommendations. As uncertainty provides a form of risk for decision makers who use boundary line outputs to make decisions, uncertainty should be included as an output of the analysis. This was also emphasized by participants in the consultation exercise in the Nairobi and Harare workshops with the majority of participants indicating that uncertainty of the outputs was important for interpretation. However, attention should not only be focused on the precise quantification of uncertainty, but should also aim to identify the sources of uncertainty and how they can be reduced (Schils et al., 2022). In this study, no measures of uncertainty were given for output of the heuristic methods because these involve subjective decisions in their implementation. Hence the uncertainty around the determined parameters will be strongly influenced by these subjective decisions, making it challenging to interpret. This mostly arises in instances where researchers have different views on reality and therefore, what to be included in a model (Walker et al., 2003).

In practice, the choice of which method to use for analysis is influenced by several factors including data availability (nature of data i.e continuous or categorical, size of data and whether it meets model assumptions among other qualities), usability of the method and the statistical competencies of the analysts (Miti et al., 2024b). For example, Smith et al. (2024) highlighted that they could not apply the cbvn method proposed by Milne et al. (2006a) on their data because the variables could not meet the assumption of normality. Currently, most published studies that have evaluated yield gaps using the boundary line methodology have used heuristic methods (Smith et al., 2024). The ease of use of the heuristic methods has been pointed out as one of the reasons for their popularity as compared to statistical methods (Miti et al., 2024b). However, the questionnaire exercise with researchers in Harare and Nairobi showed that there was no strong evidence that users found the statistical method harder to engage with, and they see value of uncertainty quantification of this approach. The cbvn provides a more robust approach to fitting boundary lines but can benefit further from improvements by extending it to include categorical variables

which currently cannot be analysed. Variables like planting date, days to weeding among others are important determinants of yield gaps and must be analysed to get a full picture of yield gaps. This was pointed out as one of the weaknesses of the cbvn by the workshop participants. Until the limitations of the cbvn method are addressed, heuristic methods will still be the preferred methods for researchers especially in agronomy where categorical variables are important factors in their studies.

The boundary line methodology is a valuable tool for extracting agronomic insights, particularly in this era of big data. However, there remains significant potential to refine this approach and broaden its applicability. Further work is needed to deepen our understanding of the agronomic implications of the boundary line methodology. While much of the current focus has been on interpretations aligned with Sprengel-Liebig Law of the Minimum (van der Ploeg et al., 1999), this perspective could be expanded to include alternative frameworks, such as the Law of the Optimum (Liebscher, 1895). Miti et al. (2024b) have outlined various ways the boundary line models can be interpreted, but more comprehensive biological exploration is required to fully understand their implications. Omission trial experiment provide a good basis to test the ability of the boundary line methodology to model the most-limiting factor. More studies with larger datasets are required to check how the detected most-limiting factor using the boundary lines compares with that from omission trials. This will enhance greater understanding on the interpretation of the boundary line methodology utilizing the Sprengel-Liebig Law of the Minimum. The experimental design of omission trials can be improved by careful thought on what variables should be measured to pick up limitations, in particular to identify limitations which a particular intervention might address (e.g. application of fertilizer). While this approach is more straightforward for less mobile nutrients, pH, and water supply (often excluded from such studies), it poses a greater challenge for dynamic variables like soil N. In omission trials, it is often assumed that adding N fertilizer addresses potential N limitations, but the complexity of soil N dynamics warrants careful consideration when interpreting results.

5. Conclusion

Our study indicates that there is consistency in the determination of the most-limiting factor amongst the different boundary line fitting methods with an overall fair to good agreement index ($\kappa > 0.4$), underscoring the coherence of these approaches for this purpose. Similarly, there is consistency in the critical soil nutrient concentration determined using different boundary line fitting methods with the exception of the binning method highlighting the impact of subjective decisions on outcomes of boundary line analysis when heuristic methods are used. These results provides a start point for the development of standard boundary line procedures for yield gap analysis to promote consistent interpretation and agronomic recommendations, and reproducible research. In contrast, our study indicates poor agreement between the boundary line methodology and both the nutrient omission experimental determination of the most-limiting factor and the RB209 established critical soil nutrient concentrations. Although there is consistency in the outputs of the different boundary line fitting methods, we recommend the use of the cbvn for determination of critical nutrient concentration as it provides more information on the uncertainty of the fitted parameters and therefore, allows for further analysis like the determination of the probability of the critical concentration falling within a given index. This is vital information for decision makers/agronomists to provide adequate solutions to close yield gaps. Furthermore, we recommend more studies with larger datasets to compare the most-limiting factors modelled by boundary line methods with those modelled experimentally to fully assess the use of boundary line methodology as a tool for determining the most-limiting factor.

CRediT authorship contribution statement

C. Miti: Writing – original draft, Visualization, Methodology, Formal analysis. **A.E. Milne:** Writing – review & editing, Supervision, Conceptualization. **K.E. Giller:** Writing – review & editing, Supervision, Conceptualization. **R.M. Lark:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Nottingham-Rothamsted Future Food Beacon Studentships in International Agricultural Development. Rothamsted Research receives strategic funding from the Biotechnology and Biological Sciences Research Council of the United Kingdom (BBSRC). We acknowledge support from the Growing Health Institute Strategic Programme [BB/X010953/1; BBS/E/RH/230003C]. We are thankful to AgSpace Agriculture Ltd, Dr. Lydia Wairegi and International maize and wheat improvement center (CIMMYT) for providing datasets 1, 2 and 3, respectively. We also acknowledge the input of the Research and Ethics committee at Sutton Bonington university of Nottingham campus for the assistance in getting the ethical approval for the workshops in Harare and Nairobi.

Appendix A. Exploratory analysis for datasets used the study

See Table A.1.

Appendix B. Evidence for upper boundary structure in data sets

See Table B.1.

Appendix C. Estimate of the standard deviation of measurement error

See Fig. C.1.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eja.2025.127744.

Data availability

• Dataset 1: A portion of this dataset is included in the BLA R package, freely available on CRAN at https://CRAN.R-project.org/package=BLA
• Dataset 2: Obtained from the study “Closing the cassava yield gap: An analysis from smallholder farms in East Africa”, published in Field Crops Research. Access to this dataset can be requested directly from the authors of that publication.
• Dataset 3 (TAMASA Ethiopia): Available on the CIMMYT Research Data & Software Repository Network (Dataverse) at https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/11015, subject to terms and conditions.

Table A.1
Summary statistics for variable in the datasets used for this study.

Data	Factor	Transformation	Power	Mean	Median	sd	Skewness	O.skewness
1	P	None	–	25.96	22	14.38	1.84	0.36
1	P	Natural log	–	3.13	3.09	0.51	0.13	0.08
1	Mg	None	–	83.60	63	68.44	4.77	0.5
1	Mg	Natural log	–	4.24	4.14	0.56	0.86	0.20
1	K	None	–	198.33	183	84.52	2.36	0.21
1	K	Natural log	–	5.21	5.21	0.38	0.16	0.001
1	pH	None	–	7.56	7.74	0.65	−0.79	−0.35
1	pH	Natural log	–	2.02	2.05	0.09	−0.99	−0.40
1	pH	Box-Cox	2	28.34	29.45	4.82	−0.63	−0.31
1	pH	Yeo-Johnson	4.99	0	0.13	1.00	−0.27	−0.19
1	Yield	None	–	9.25	9.36	1.85	−0.48	−0.06
2	pH	None	–	5.82	5.9	0.46	−0.04	−0.15
2	SOC	None	–	1.72	1.61	0.63	1.11	0.32
2	SOC	Natural log	–	0.49	0.48	0.35	0.03	0.12
2	N	None	–	0.083	0.077	0.044	0.80	0.33
2	N	Natural log	–	−2.63	−2.56	0.60	−0.89	0.02
2	P	None	–	4.93	4.03	3.25	2.35	0.40
2	P	Natural log	–	1.44	1.39	0.53	0.56	0.13
2	K	None	–	0.46	0.39	0.31	1.56	0.37
2	K	Natural log	–	−0.98	−0.93	0.65	−0.03	0.01
2	Ca	None	–	2.76	1.63	2.79	0.99	0.55
2	Ca	Natural log	–	0.29	0.49	1.40	−0.43	−0.18
2	Mg	None	–	1.11	0.79	0.99	0.98	0.51
2	Mg	Natural log	–	−0.41	−0.24	1.18	−0.80	−0.19
2	Yield	None	–	13.97	13.86	6.32	0.15	0.013
3	pH	None	–	5.39	5.4	0.40	0.19	0.11
3	OC	None	–	2.17	2.16	0.46	−0.06	0.08
3	N	None	–	0.22	0.23	0.05	−0.07	−0.14
3	P	None	–	5.04	3.3	4.91	2.74	0.49
3	P	Natural log	–	1.25	1.19	0.87	−0.01	0.04
3	Ca	None	–	7.77	7.4	2.25	0.65	0.28
3	Ca	Natural log	–	2.01	2.00	0.29	−0.01	0.14
3	Mg	None	–	3.24	3.2	1.01	0.96	0.07
3	K	None	–	0.98	0.99	0.41	1.12	−0.01
3	Yield	None	–	3.34	2.81	2.09	0.82	0.29

Table B.1
The probability of observing a larger peel clustering (lower *sd* of euclidean distance of boundary points to centre) than that of a normal bivariate joint distribution on the left (*l*) and right (*r*) sections.

Dataset	n	Variable	sd_l	$\tilde{sd}_l$	p-value $_l$	sd_r	$\tilde{sd}_r$	p-value $_r$
1	6010	P	1.045	1.181	0.019	1.115	1.276	0.013
1	6010	Mg	1.309	1.153	0.997	1.087	1.250	0.008
1	6010	pH	0.952	1.064	0.035	1.244	1.155	0.877
2	110	SOC	3.610	3.643	0.484	4.077	4.115	0.482
2	110	N	2.778	3.567	0.09	3.990	4.091	0.459
2	110	pH	3.186	3.626	0.238	3.985	4.106	0.423
2	110	P	3.345	3.571	0.369	4.036	4.070	0.505
2	110	K	2.837	3.556	0.111	4.023	3.985	0.567
2	110	Mg	3.2603	3.408	0.421	3.781	3.838	0.494
2	110	Ca	3.025	3.469	0.244	4.330	3.736	0.830
3	148	N	1.169	1.358	0.169	1.541	1.417	0.742
3	148	pH	1.410	1.135	0.841	1.506	1.411	0.679
3	148	SOC	1.151	1.246	0.379	1.530	1.358	0.746
3	148	P	1.330	1.081	0.830	1.382	1.244	0.739
3	148	Ca	1.464	1.206	0.837	1.496	1.409	0.658
3	148	Mg	1.344	1.043	0.869	1.430	1.213	0.816
3	148	K	1.499	1.264	0.802	1.408	1.387	0.553

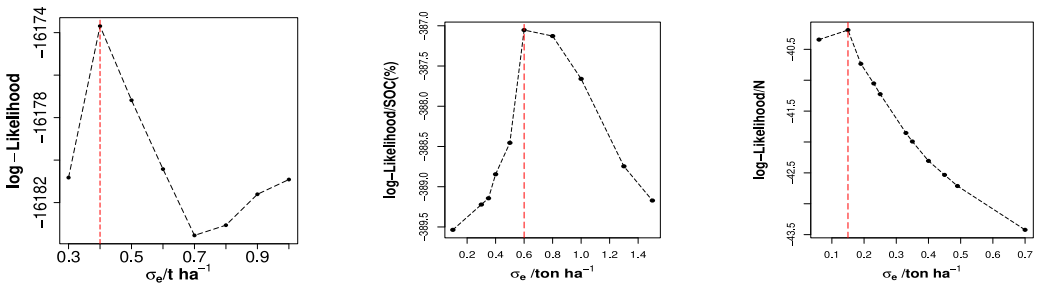

Fig. C.1. Estimate of σ_{me} using the log-likelihood profile for (a) data set 1, (b) Dataset 2 and (c) Dataset 3. The red dashed line represents the estimate of σ_{me}.

References

AHDB, 2023. Nutrient Management Guide (RB209). Section 1: Arable Crops. Agriculture and Horticulture Development Board, Coventry, UK.

Andrade, C.B., Comin, J.J., Moura-Bueno, J.M., Brunetto, G., 2023. Obtaining reference values for nutrients in vineyard soils through boundary line approach using Bayesian segmented quantile regression on commercial farm data. Eur. J. Agron. 150, 126928.

Bargagli Stoffi, F.J., Cevolani, G., Gnecco, G., 2022. Simple models in complex worlds: Occam's razor and statistical learning theory. Minds Mach. 32 (1), 13–42.

Barry, J., 2011. Doing Bayesian data analysis: A tutorial with R and BUGS. Europe' s J. Psychol. 7 (4), 778.

Baudron, F., Ndoli, A., Habarurema, I., Silva, J.V., 2019. How to increase the productivity and profitability of smallholder rainfed wheat in the Eastern African highlands? Northern Rwanda as a case study. Field Crop. Res. 236, 121–131.

Brys, G., Hubert, M., Struyf, A., 2008. Goodness-of-fit tests based on a robust measure of skewness. Comput. Statist. 23, 429–442.

Burnham, K., Anderson, D., 2004. Multimodel inference, understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.

Cade, B.S., Terrell, J.W., Schroeder, R.L., 1999. Estimating effects of limiting factors with regression quantiles. Ecology 80, 311–323.

Casanova, D., Goudriaan, J., Bouma, J., Epema, G., 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice. Geoderma 91 (3–4), 191–216.

Craufurd, P., 2017. TAMASA Ethiopia. Nutrient Omission Trial (NOT) Datasets for 2015 and 2016 Seasons. CIMMYT Research Data & Software Repository Network.

Dobson, A.J., Barnett, A.G., 2018. An Introduction to Generalized Linear Models. CRC Press.

Draper, N.R., Smith, H., 1998. Applied Regression Analysis, vol. 326, John Wiley & Sons.

Evanylo, G.K., Sumner, M.E., Taylor & Francis, 1987. Utilization of the boundary line approach in the development of soil nutrient norms for soybean production. Commun. Soil Sci. Plant Anal. 18, 1397–1401.

FAO, DWFI, 2015. Yield gap analysis of field crops: Methods and case studies, by Sadras, V.O.,Cassman, K.G.G.,Grassini, P., Hall, A.J., Bastiaanssen, W.G.M., Laborte, A.G., Milne, A.E., Sileshi, G., Steduto, P. FAO Water Rep. 41.

Fermont, A.M., Van Asten, P.J., Tittonell, P., Van Wijk, M.T., Giller, K.E., 2009. Closing the cassava yield gap: an analysis from smallholder farms in East Africa. Field Crop. Res. 112 (1), 24–36.

Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76 (5), 378.

Fleiss, J.L., Cohen, J., Everitt, B.S., 1969. Large sample standard errors of kappa and weighted kappa. Psychol. Bull. 72 (5), 323.

Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical methods for rates and proportions. In: Statistical Methods for Rates and Proportions, third ed. In: Wiley series in probability and statistics, Wiley-Interscience, Hoboken, N.J.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., et al., 2011. Solutions for a cultivated planet. Nature 478 (7369), 337–342.

Gamer, M., Lemon, J., Gamer, M.M., Robinson, A., Kendall's, W., 2012. Package 'irr'. Var. Coefficients Interrater Reliab. Agreem. 22, 1–32.

Giller, K.E., Delaune, T., Silva, J.V., Descheemaeker, K., van de Ven, G., Schut, A.G., van Wijk, M., Hammond, J., Hochman, Z., Taulya, G., Chikowo, R., Narayanan, S., Kishore, A., Bresciani, F., Teixeira, H.M., Andersson, J.A., van Ittersum, M.K., 2021. The future of farming: Who will produce our food? Food Secur. 13, 1073–1099.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: The challenge of feeding 9 billion people. Science 327 (5967), 812–818.

Hajjarpoor, A., Soltani, A., Zeinali, E., Kashiri, H., Aynehband, A., Vadez, V., 2018. Using boundary line analysis to assess the on-farm crop yield gap of wheat. Field Crop. Res. 225, 64–73.

Harris, E.F., Smith, R.N., 2009. Accounting for measurement error: A critical but often overlooked process. Arch. Oral. Biol. 54.

Helidoniotis, F., Haddon, M., Tuck, G., Tarbath, D., 2011. The relative suitability of the von Bertalanffy, Gompertz and inverse logistic models for describing growth in blacklip abalone populations (Haliotis rubra) in Tasmania, Australia. Fish. Res. 112 (1–2), 13–21.

Koenker, R., D'Orey, V., 1987. Computing regression quantiles, applied statistics. Appl. Stat. 36, 383–393.

Kopittke, P.M., Menzies, N.W., Wang, P., McKenna, B.A., Lombi, E., 2019. Soil and the intensification of agriculture for global food security. Environ. Int. 132, 105078.

Lark, R.M., Gillingham, V., Langton, D., Marchant, B.P., 2020. Boundary line models for soil nutrient concentrations and wheat yield in national-scale datasets. Eur. J. Soil Sci. 71, 334–351.

Lark, R.M., Milne, A.E., 2016. Boundary line analysis of the effect of water-filled pore space on nitrous oxide emission from cores of arable soil. Eur. J. Soil Sci. 67, 148–159.

Lee, P.H., Yu, P.L., 2013. An R package for analyzing and modeling ranking data. BMC Med. Res. Methodol. 13, 1–11.

Liebscher, G., 1895. Untersuchungen über die Bestimmung des Düngerbedürfnisses der Ackerböden und Kulturpflanzen. J. Landwirtsch. 43, 49–125.

Makowski, D., Doré, T., Monod, H., 2007. A new method to analyse relationships between yield components with boundary lines. Agron. Sustain. Dev. 27, 119–128.

Marden, J.I., 1996. Analyzing and Modeling Rank Data. CRC Press.

Meredith, M., Kruschke, J., 2020. HDInterval: Highest (posterior) Density Intervals. R package version 0.2.

Milne, A.E., Ferguson, R.B., Lark, R.M., 2006a. Estimating a boundary line model for a biological response by maximum likelihood. Ann. Appl. Biol. 149, 223–234.

Milne, A.E., Wheeler, H.C., Lark, R.M., 2006b. On testing biological data for the presence of a boundary. Ann. Appl. Biol. 149, 213–222.

Miti, C., Lark, R.M., Milne, A.E., Giller, K.E., 2024a. BLA: Boundary line analysis. R package version 1.0.1.9000.

Miti, C., Milne, A., Giller, K., Lark, R., 2024b. The concepts and quantification of yield gap using boundary lines. A review. Field Crop. Res. 311, 109365.

Miti, C., Milne, A., Giller, K., Sadras, V., Lark, R., 2024c. Exploration of data for analysis using boundary line methodology. Comput. Electron. Agric. 219, 108794.

Mitscherlich, E., 1909. Des Gesetz des Minimums und das Gesetz des abnehmended Bodenertrages. Landwirsch Jahrb 3, 537–552.

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. Ecol. Model. 62 (4), 275–293.

Myers, R.H., Myers, R.H., 1990. Classical and Modern Regression with Applications, vol. 2, Duxbury press Belmont, CA.

Nelder, J., 1961. The fitting of a generalization of the logistic curve. Biometrics 17 (1), 89–110.

Oliver, F., 1964. Methods of estimating the logistic growth function. J. R. Stat. Soc. Ser. C. Appl. Stat. 13 (2), 57–66.

PDA, 2017. Magnesium as a Nutrient for Crops and Grass. Potash Development Association, https://www.pda.org.uk/magnesium-nutrient-crops-grass/. (Accessed 29 April 2024).

Pruthviraj, N., Murali, K., Chaitanya, A., Harish, M., Karthik, A., 2024. Exploring the dynamics of nitrogen from conventional manures in the soil plant atmosphere continuum: A comprehensive review. Commun. Soil Sci. Plant Anal. 55 (11), 1690–1701.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M.B., 2013. Package 'MASS'. CRAN 538, 113–120.

Rousseeuw, P.J., Ruts, I., Tukey, J.W., 1999. The bagplot: a bivariate boxplot. Amer. Statist. 53 (4), 382–387.

Royall, R., 2017. Statistical Evidence: A Likelihood Paradigm. Routledge.

Sadras, V.O., 2020. On water-use efficiency, boundary functions, and yield gaps: French and Schultz insight and legacy. Crop. Sci. 60, 2187–2191.

Schils, R.L., van Voorn, G.A., Grassini, P., van Ittersum, M.K., 2022. Uncertainty is more than a number or colour: Involving experts in uncertainty assessments of yield gaps. Agricult. Sys. 195, 103311.

Schmidt, U., Thöni, H., Kaupenjohann, M., 2000. Using a boundary line approach to analyze N_2O flux data from agricultural soils. Nutr. Cycl. Agroecosystems 57, 119–129.

Schnug, E., Heym, J.M., Murphy, D.P.L., 1995. Boundary line determination technique (BOLIDES). In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Site-Specific Management for Agricultural Systems. Wiley Online Library, pp. 899–908.

Schut, A.G., Giller, K.E., 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. Geoderma 380, 114680.

Shabani, A., Sepaskhah, A., Kamgar-Haghighi, A., Honar, T., 2018. Using double logistic equation to describe the growth of winter rapeseed. J. Agric. Sci. 156 (1), 37–45.

Shatar, T.M., McBratney, A.B., 2004. Boundary-line analysis of field-scale yield response to soil properties. J. Agric. Sci. 142, 553–560.

Smith, H.W., Ashworth, A.J., Nalley, L.L., Schmidt, A., Turmel, M.-S., Bucaro, A., Owens, P.R., 2024. Boundary line analysis and machine learning models to identify critical soil values for major crops in Guatemala. Agron. J. 116 (3), 1071–1087.

Tang, G., Shafer, S.L., Bartlein, P.J., Holman, J.O., 2009. Effects of experimental protocol on global vegetation model accuracy: a comparison of simulated and observed vegetation patterns for Asia. Ecol. Model. 220 (12), 1481–1491.

Tarkesh, M., Jetschke, G., 2012. Comparison of six correlative models in predictive vegetation mapping on a local scale. Environ. Ecol. Stat. 19 (3), 437–457.

Tukey, J.W., 1975. Mathematics and the picturing of data. In: Proceedings of the International Congress of Mathematicians, vol. 2, Vancouver, pp. 523–531.

van der Ploeg, R.R., Böhm, W., Kirkham, M.B., 1999. On the origin of the theory of mineral nutrition of plants and the law of the minimum. Soil Sci. Am. J. 63 (5), 1055–1062.

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer.

Walker, W.E., Harremoës, P., Rotmans, J., Van Der Sluijs, J.P., Van Asselt, M.B., Janssen, P., Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr. Assess. 4 (1), 5–17.

Wang, N., Jassogne, L., van Asten, P.J., Mukasa, D., Wanyama, I., Kagezi, G., Giller, K.E., 2015. Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. Eur. J. Agron. 63, 1–11.

Webb, R.A., 1972. Use of the boundary line in the analysis of biological data. J. Hortic. Sci. 47 (3), 309–319.

Yeo, I.-K., Johnson, R.A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87 (4), 954–959.