Yield Gap Analysis Using the Boundary-line Methodology

Chawezi Miti

University of Nottingham/Rothamsted Research

April 23, 2025

Course Overview

This course is grounded in the boundary line methodology introduced by Webb (1972). This approach is used to analyze biological data collected in uncontrolled (non-experimental) environments, aiming to identify the most efficient biological response. It has been widely applied in environmental research.
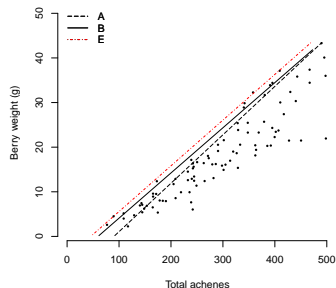
Course Objectives

The objective of the course are to:

1. Understand the principle behind the boundary-line methodology.

2. Apply the boundary line methodology to data using the BLA R package.

3. Make agronomic interpretations from outputs of a boundary line analysis

Course Outcomes

At the end of the course, you should be able to:

1. Identify the conditions under which boundary line analysis is appropriate.

2. Fit boundary line models to data using various methods.

3. Interpret the results of boundary line analysis in an agronomic context

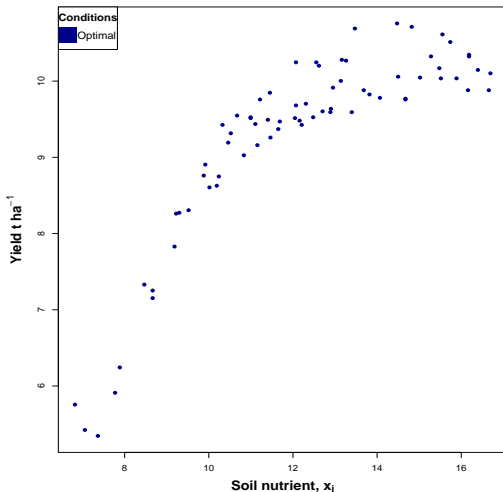## Course Content

The course is composed of two parts:

shutterstock.com · 1586195833

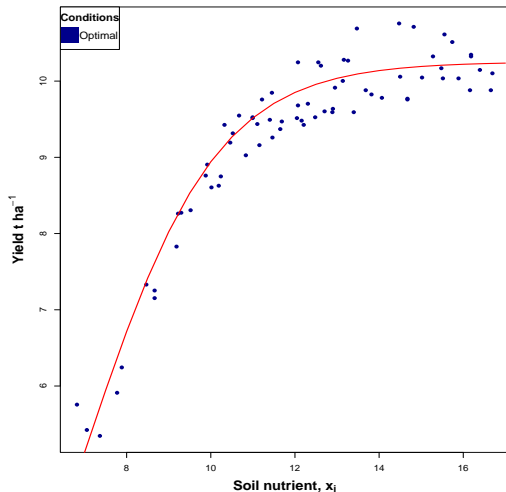**What are boundary-line Models?**

The Boundary Line Methodology
Boundary line analysis

- Data collected in controlled condition
- All conditions are optimal except x_i
- Maximum biological limit

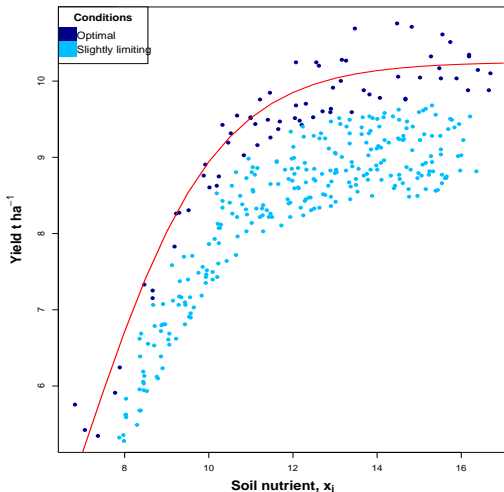# The Boundary Line Methodology
Boundary line analysis

- A model can be fitted to these data
- $yield = f(x_i) + e$
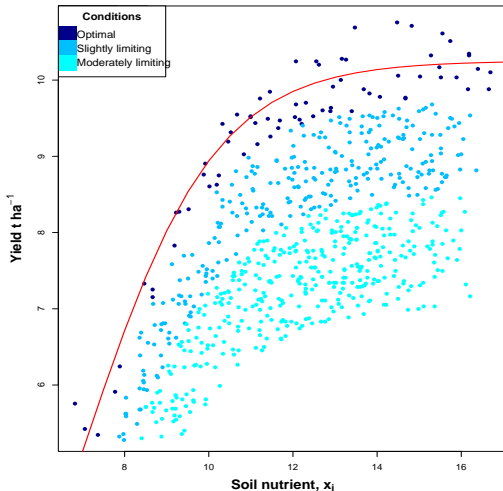- Least squares method

The Boundary Line Methodology

Boundary line analysis

- Additional data from conditions
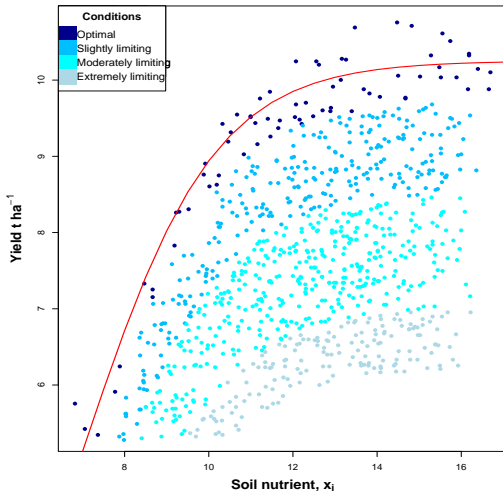 1. Slightly limiting

The Boundary Line Methodology

Boundary line analysis

- Additional data from conditions
  1. Slightly limiting
  2. Moderately limiting
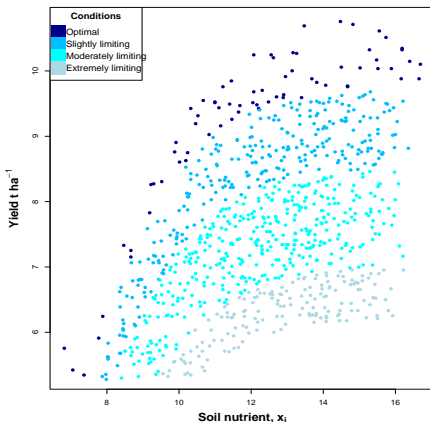
The Boundary Line Methodology

Boundary line analysis

- Additional data from conditions
  1. Slightly limiting
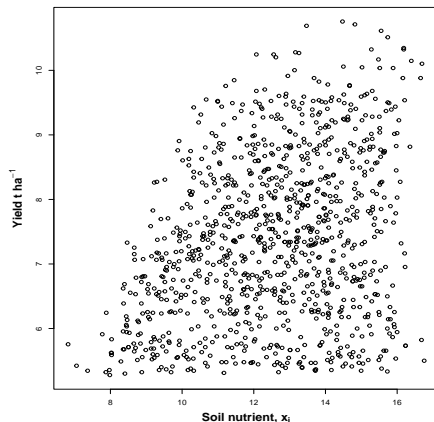 2. Moderately limiting
  3. Extremely limiting

The Boundary Line Methodology
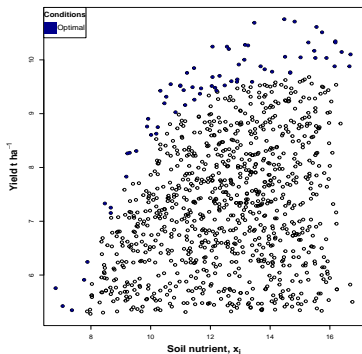
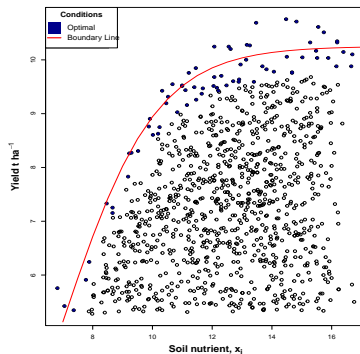Boundary line analysis

- Conceptually

- Reality

The Boundary Line Methodology
Boundary line analysis

Webb(1972) proposed an upper boundary model for such data
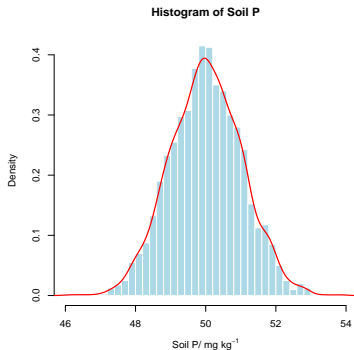

- Identify data in optimal conditions

- Fit a boundary model to these data

What conditions are necessary to apply Boundary-line analysis?

Conditions to apply Boundary Line Analysis
Assumptions for boundary line modelling

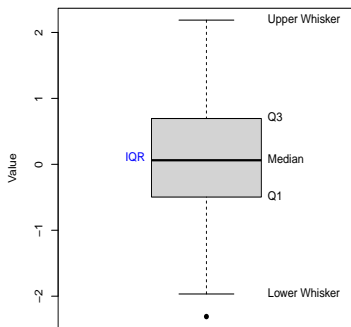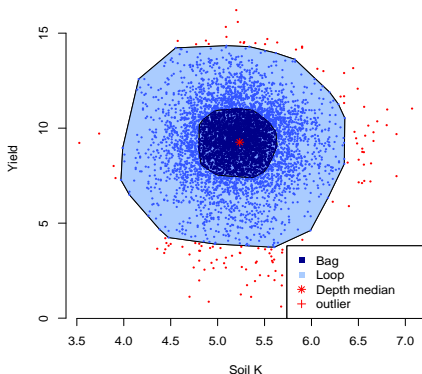## 1. Normal distribution for x and y variables

Histogram of Soil P

- Variable should cover a wide range of instances
- Skewness [-1,1]
- Octile-skewness [-0.2,0.2]

Conditions to apply Boundary Line Analysis

Assumptions for boundary line modelling

2. Outlier detection and removal
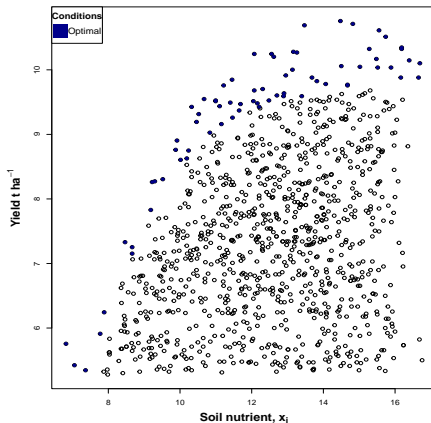- Use of bagplot, a bivariate equivalent of univariate boxplot

Conditions to apply Boundary Line Analysis
Assumptions for boundary line modelling

3. Evidence of boundary conditions
- Large number of data points in the boundary neighbourhood
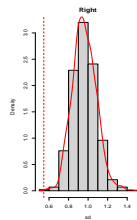
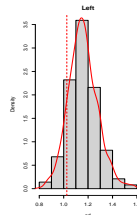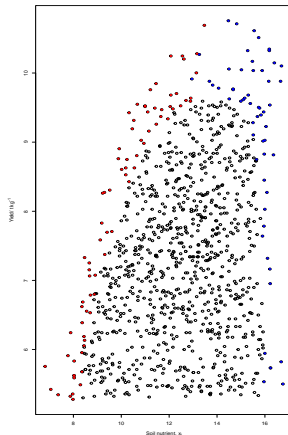# Conditions to apply Boundary Line Analysis
Assumptions for boundary line modelling

3. **Evidence of boundary limiting conditions**
- There should be evidence of data clustering at upper edges
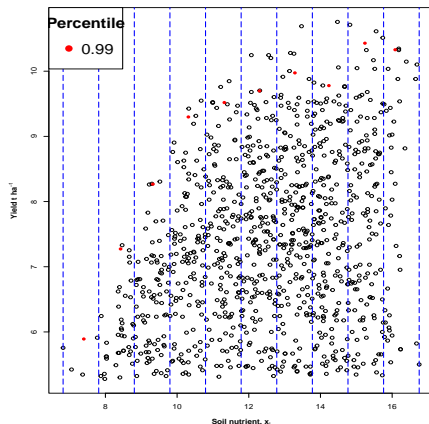

- Indices
  - Area
  - Distance

How is a boundary line model fitted?

Model fitting
Methods of Boundary line fitting

1. Binning method
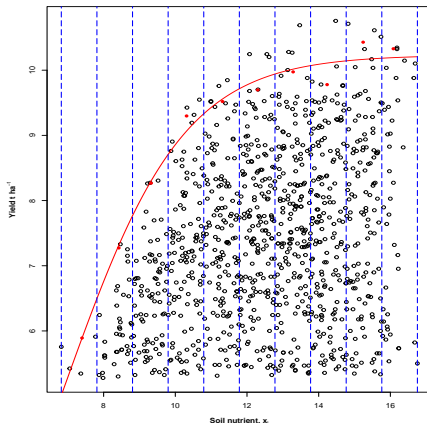

- Divide the data into sections
- Select boundary points
- Fit boundary model (OLS)

Model fitting
Methods of Boundary line fitting

1. Binning method

- Identify a biologically plausible model
- Fit boundary model
- Ordinary least squares

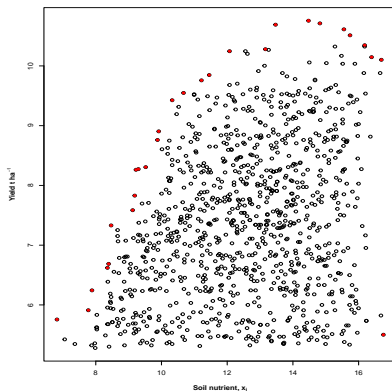$$\hat{\beta}_\tau = \min \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2$$

Model fitting
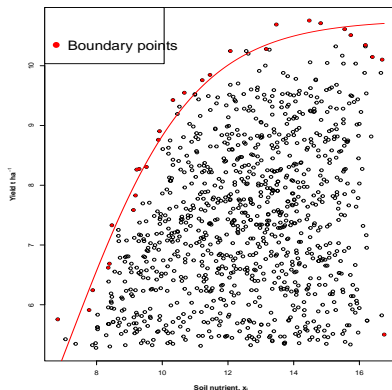## Methods of Boundary line fitting

2. Bolides algorithm
a) Select boundary points
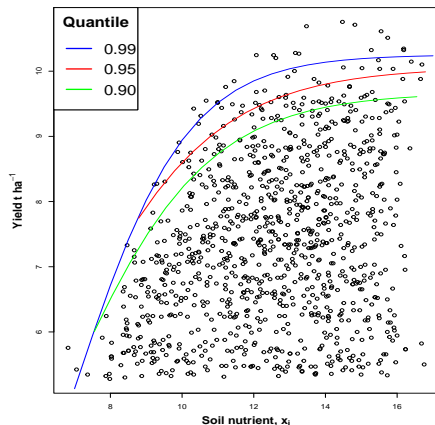
- Points that enclose data

b) Fit boundary model

- Ordinary least squares

Model fitting
Methods of Boundary line fitting

3. Quantile regression

- Fit boundary models based on a quantile value (τ)

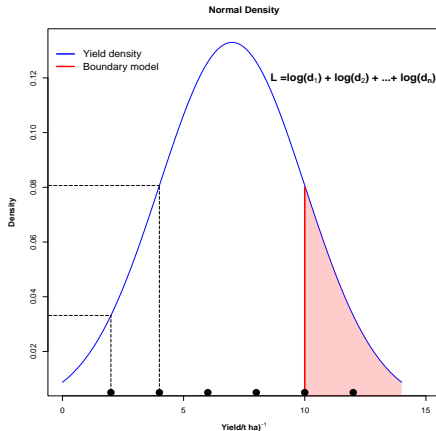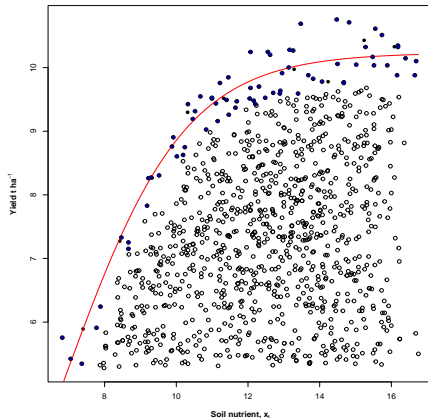$$\hat{\beta}_\tau = \min \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta)$$

- β is model parameter(s)
- $+$ residues, weight $= \tau$
- $-$ residues, weight $= 1 - \tau$

Model fitting
Methods of Boundary line fitting

4. Censored bivariate normal model
- Based on clustering structure (ML)

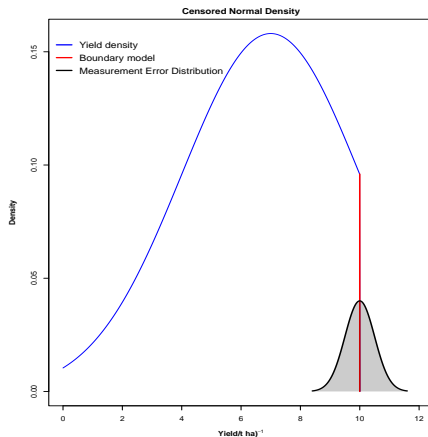# Model fitting
Methods of Boundary line fitting

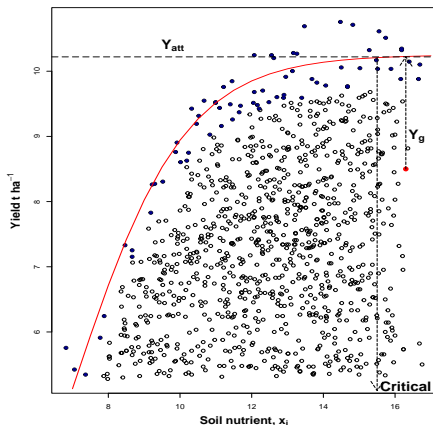## 4. Censored bivariate normal model
- Based on clustering structure and measurement error (ML)

How are boundary-line models interpretation agronomically?

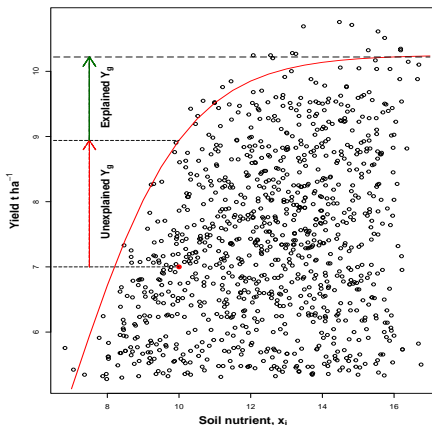# Agronomic Interpretation of Boundary Line models

1. Yield gap and Critical nutrient values

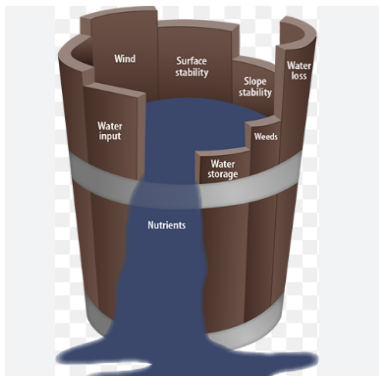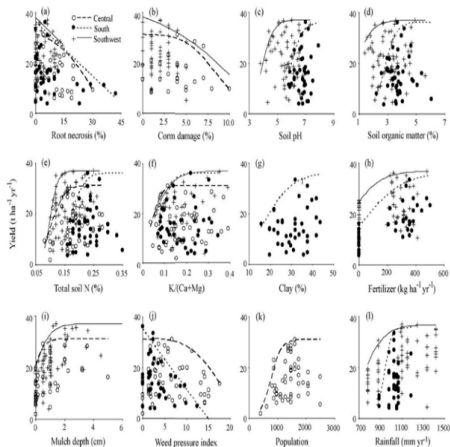# Agronomic Interpretation of Boundary Line models

2. **Explained and Unexplained yield gaps**
- Uni-factor analysis

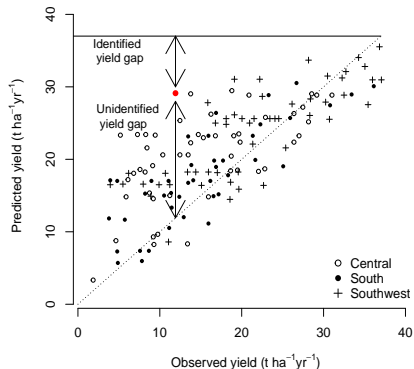# Agronomic Interpretation of Boundary Line models

3. Most limiting factor (Liebig, 1840)

Agronomic Interpretation of Boundary Line models

4. Explained and Unexplained yield gaps

- Multi-factor analysis

What tools are available to carry-out a boundary-line analysis?

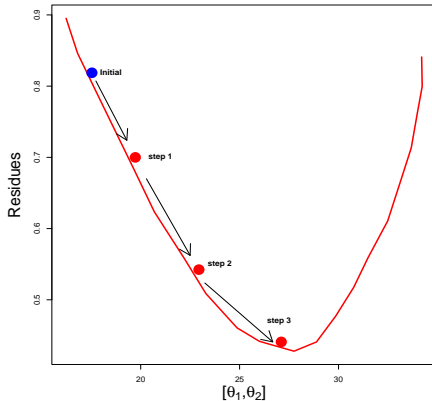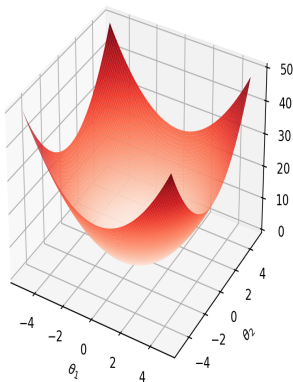# Applications
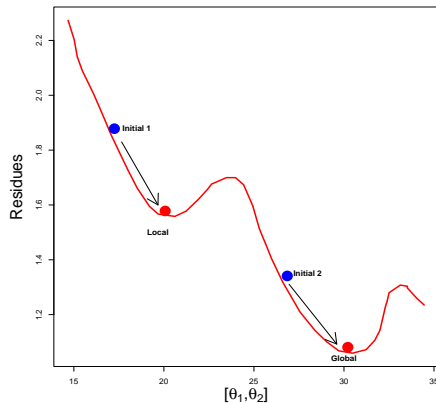
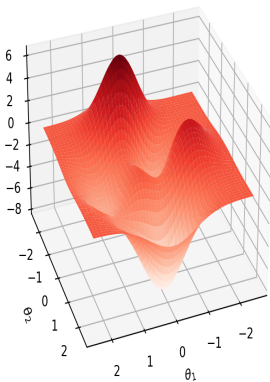## BLA R Package

Practical in R

Optimization
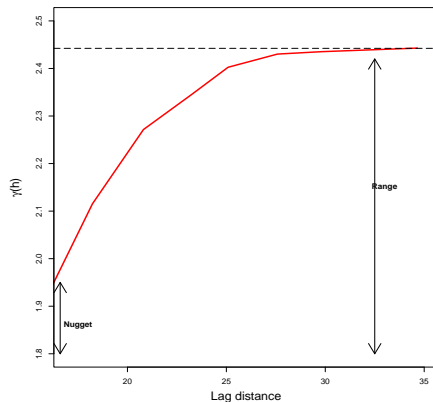
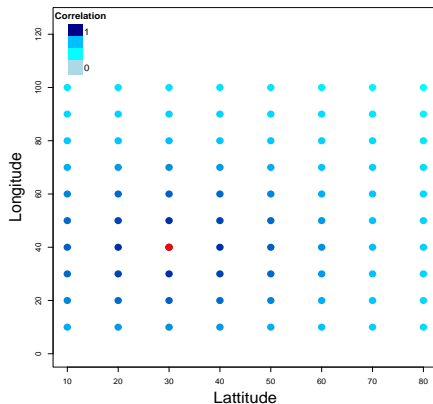- **Initial-starting** values for parameter optimization

Optimization

- **Local** and **Global** optima in parameter optimization

Measurement error determination for cbvn

① Direct measurement
② Can be determined from **Nugget** variance of a **variogram**

Measurement error determination for cbvn

1. Can be determined from **Likelihood** profile of σ_{me}