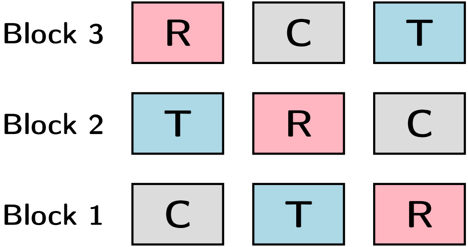
Geomatic Techniques to Support Phytosanitary Products Tests whithin the EPPO Standard Framework

Samuele Bumbaca

University of Turin

August 28, 2025

The Traditional Approach to Agricultural Trials



C Control
T Tested Product
R Reference Product

ANOVA Model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Where:

- y_{ij} = response
- μ = overall mean
- α_i = treatment effect
- β_j = block effect
- ε_{ij} = random error

Note:

This is the additive model. Modern approaches may include interaction terms: $\alpha_i \times \beta_j$

Key Assumptions of Traditional ANOVA

Statistical Assumptions:

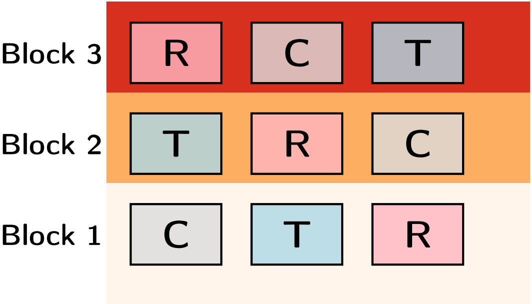
- **Randomization**: Treatments randomly assigned within blocks
- **Replication**: Each treatment appears in each block
- **Independence**: Observations are independent given the design
- **Homoscedasticity** : Equal variances across treatments
- **Normality**: Residuals follow normal distribution

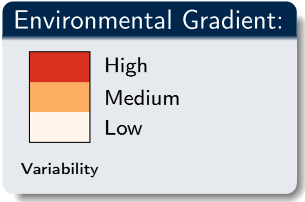
Consequences of Assumption Violations:

- **Invalid conclusions of parametric tests**: Need for non-parametric tests leading to reduced statistical power

Based on R. A. Fisher, Statistical Methods for Research Workers, in S. Kotz & N. L. Johnson (eds.), Breakthroughs in Statistics: Methodology and Distribution, pp. 66–70, Springer, New York, 1992.

The Right Blocking: Capturing Environmental Variability



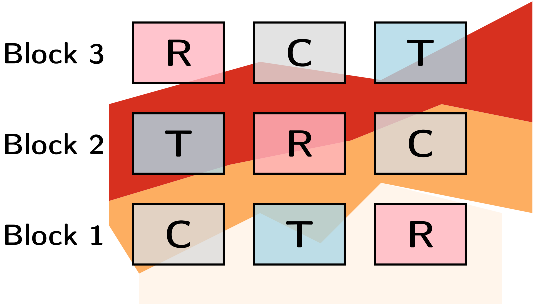


C Control
T Tested Product
R Reference Product

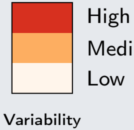
Success of Blocking Strategy:

- **Within-block homogeneity**: Treatments compared under similar conditions
- **Between-block heterogeneity**: Environmental gradient captured by block effects

The Wrong Blocking: Assumption Violation



Environmental Gradient:



C	Control
T	Tested Product
R	Reference Product

Heteroscedasticity Assumption Violation Problem:

- **Blocks fail to capture environmental variability**: Treatments compared under different conditions
- **Invalid parametric test**: Residual variance differs across treatments

Current Limitations in Statistics for Agricultural Trials

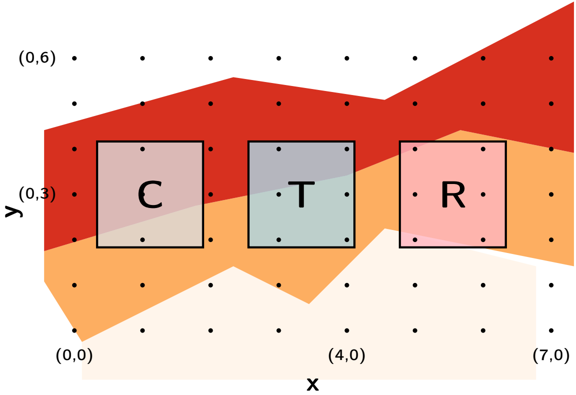
Traditional Approach Issues:

- **Human-dependent blocking**: Environmental variability assessment relies on experimenter experience
- **A priori identification**: Must identify variance sources BEFORE data collection

The Challenge:

How can we capture environmental variability mathematically rather than through human judgment?

Geostatistical Approach: Spatial Linear Mixed Models



C/T/R Control/Tested/Reference
• Georeferenced observations

Spatial LMM:

$$y(s_i) = \mu + \alpha_j + f(s_i) + \varepsilon_i$$

Where:

- $y(s_i)$ = response at s_i
- μ = overall mean
- α_j = treatment effect
- $f(s_i)$ = spatial random field
- ε_i = error
- $s_i = (x_i, y_i)$ = coordinates

Benefits:

- **No blocking**: Spatial correlation captures variability
- **Post-hoc**: No a priori variance identification
- **Homoscedasticity**: Assumption satisfied in more cases in respect blocking

The Missing Link: Spatial Coordinates

Geostatistical Methods Advantages:

- ✓ **Mathematical modeling** of environmental variability
- ✓ **Post-hoc analysis** - no need for prior knowledge of the environment variables and of their distribution
- ✓ **Superior performance** in handling spatial heterogeneity
- ✓ **EPPO recognized** approach

Current Barrier:

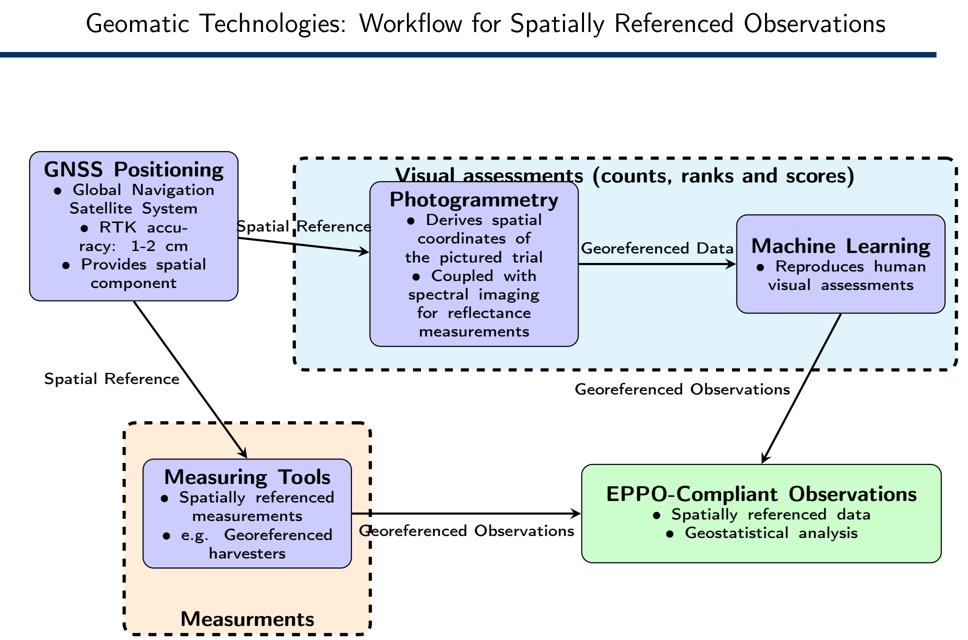
- × **Requires spatially referenced observations**
- × **Traditional manual assessments lack coordinates**
- × **Implementation gap** in practical field trials

Central Research Question

Can geomatics technologies provide spatially referenced observations that enable geostatistical analysis within EPPO-compliant Plant Protection Product trials?

Specific Objectives:

1. Establish which geomatics technologies can be used to collect spatially referenced observations
2. Demonstrate the feasibility of collect spatially referenced observations in compliant with EPPO standards
3. Validate performance against traditional methods
4. Provide practical implementation guidelines


Geomatic Technologies: Workflow for Spatially Referenced Observations
GNSS Positioning
• Global Navigation Satellite System
• RTK accuracy: 1-2 cm
• Provides spatial component
Spatial Reference
Visual assessments (counts, ranks and scores)
Photogrammetry
• Derives spatial coordinates of the pictured trial
• Coupled with spectral imaging for reflectance measurements
Georeferenced Data
Machine Learning
• Reproduces human visual assessments
Spatial Reference
Georeferenced Observations
Measuring Tools
• Spatially referenced measurements
• e.g. Georeferenced harvesters
Georeferenced Observations
EPPO-Compliant Observations
• Spatially referenced data
• Geostatistical analysis
Measurments

Georeferencing EPPO Standard Assessments

Table: Different modes of observation and types of variables

Type of Variable	Measurement	Ranking	Scoring
Binary			X
Nominal			X
Ordinal		X	X
Discrete	X		
Continuous limited	X		
Continuous not limited	X		

Summary from EPPO PP 1/152: Design and analysis of efficacy evaluation trials

Current State of Georeferencing in Agricultural Trials:

EPPO's continuous, unbounded measurements are typically tool-collected and easily georeferenced (e.g., yield harvesters), whereas other regulated variables depend on experimenters' visual assessments, complicating spatial integration.

Statistical Analysis Methods in Field Trials

Traditional Approach

Randomized Complete Block Design (RCBD)

- Assumes blocks capture all spatial variation
- Fixed block effects
- Cannot model continuous spatial patterns
- Residual spatial structure ignored

Geostatistical Approach

Spatial Analysis Methods

- Model spatial correlation explicitly
- Continuous spatial trends
- Better residual structure
- Improved precision

Research Question

Can geostatistical methods provide better estimates when environmental variation is not perfectly captured by experimental blocks?

Simulated Trial Design

Experimental Setup

- 3 treatments $\times$ 3 blocks (9 plots)
- 15m $\times$ 10m plots with 17 measurement points each
- **Spatial gradient:** -1.5 to +1.5 t/ha across field
- **Treatment effects:** Control (0), Test (+2), Reference (+1) t/ha

Block 1	Block 2	Block 3
Test	Control	Reference
Reference	Test	Control
Control	Reference	Test

Table: Randomization Layout

Spatial Pattern

Environmental gradient:
West $\rightarrow$ East: 10.3 $\rightarrow$ 14.2 t/ha

Key Issue

Blocks are not perfectly aligned with environmental gradient, creating spatial confounding

RCBD Analysis Results

Model Specification

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

- μ: overall mean
- τ_i: treatment effect
- β_j: block effect
- ε_{ij}: random error

ANOVA Table

Source	DF	MS	F	P-value
Treatment	2	5.034	316.2	$<$0.001
Block	2	0.623	39.1	0.002
Error	4	0.016	–	–

Model Performance

- $R^2 = 0.994$
- Residual SE = 0.126
- Spatial structure in residuals ignored

Treatment Estimates

Treatment	Effect	SE
Control	0.00	–
Reference	+2.03	0.089
Test	+2.41	0.089

Limitation

Variogram Analysis

Geostatistical Approach

- Model spatial correlation explicitly
- Variogram: $\gamma(h) = \frac{1}{2}E[(Z(s) - Z(s+h))^2]$
- Fitted model: Linear variogram
- Parameters:
  - Nugget: 0.000
  - Sill: 0.121
  - Range: 1.35m

Spatial Model

$$Y(s) = \mu + X(s)\beta + Z(s)$$

where $Z(s)$ follows spatial covariance

Advantages over RCBD

- Continuous spatial modeling
- Better prediction at unsampled locations
- Accounts for spatial autocorrelation
- More efficient parameter estimation

Treatment Effects

Similar estimates to RCBD but with:

- Spatial correction applied
- Reduced standard errors
- Better residual structure

P-Splines Analysis (SpATS)

Spatial Splines Model

- Smooth spatial trends using P-splines
- Flexible non-parametric approach
- Automatic smoothing parameter selection
- Handles complex spatial patterns

Model Specification

$$Y = X\beta + f(x, y) + \varepsilon$$

- $f(x, y)$: smooth spatial surface
- Penalized B-splines basis

Model Results

- Spatial variance explained: 6.8%
- Error variance: 0.068
- Effective dimensions:
  - f(x,y)|x: 0.790
  - f(x,y)|y: 1.344
- Deviance: -231.86

Interpretation

- Moderate spatial pattern detected
- Stronger trend in Y direction
- Complements RCBD block structure

Methods Comparison Summary

Method	Spatial Modeling	Flexibility	Assumptions
RCBD	Discrete blocks	Low	Blocks capture all variation
Variogram	Continuous correlation	Medium	Stationary covariance
P-Splines (SpATS)	Smooth surfaces	High	Minimal assumptions

When RCBD Fails

- Blocks don't align with gradients
- Complex spatial patterns
- Residual spatial autocorrelation
- Underestimated treatment precision

Geostatistical Benefits

- Model true spatial structure
- Improved parameter estimation
- Better experimental precision
- Spatial prediction capability

Conclusion

Geostatistical methods provide superior analysis when environmental variation exceeds the capacity of experimental blocking

Practical Implications for PPP Trials

Current Practice

- RCBD widely used in regulatory trials
- Fixed blocking strategies
- Spatial information often ignored
- Conservative approach to meet EPPO standards

Recommended Approach

- Collect spatial coordinates
- Use RCBD as baseline
- Apply geostatistical diagnostics
- Consider spatial methods when:
  - Residual spatial correlation
  - Complex field gradients
  - Precision requirements

Regulatory Requirements

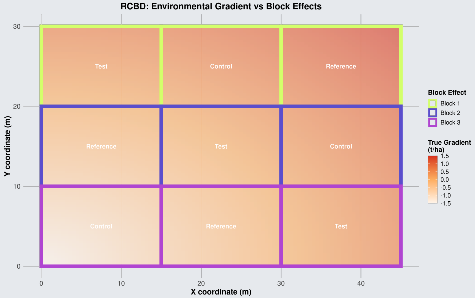
- $R^2 > 0.85$ for model acceptance
- All methods achieved this threshold
- Focus on treatment effect precision

Future Directions

- Integration in regulatory guidelines
- Automated spatial analysis tools
- Training for practitioners

RCBD: Trial Design and Block Effects

Trial Design with Environmental Gradient



Blue: True environmental gradient Red: Estimated block effects

Model Formula

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

μ : Overall mean

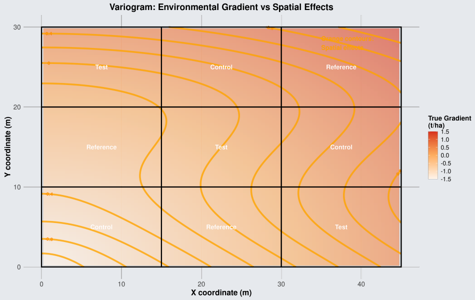
τ_i : Treatment effect

β_j : Block effect

ε_{ij} : Random error

Variogram: Trial Design and Spatial Effects

Trial Design with Environmental Gradient



Blue: True environmental gradient Orange: Estimated spatial effect

Model Formula

$$Y(s) = \mu + X(s)\beta + Z(s)$$

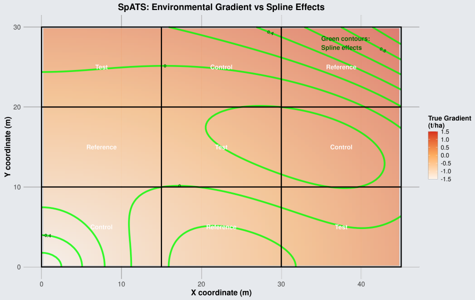
μ : Overall mean

$X(s)\beta$: Treatment effect

$Z(s)$: Spatial effect (covariance)

SpATS: Trial Design and Spline Effects

Trial Design with Environmental Gradient



Blue: True environmental gradient Green: Estimated spline spatial effect

Model Formula

$$Y = X\beta + f(x, y) + \varepsilon$$

$X\beta$: Treatment effect

$f(x, y)$: Smooth spatial surface

ε : Random error

Summary of Estimated Effects

Model	Treatment Effect (Test)	Treatment Effect (Reference)	Environment
RCBD	+2.41	+2.03	Block (0.47
Variogram	+2.41	+2.03	Spatial (Sill: 0.121,
SpATS	+2.41	+2.03	Spline (6.8%

Table: Comparison of estimated effects for each model