SAGEA
Group

DISAFA

UNIVERS
DI TORI

Geomatics Technologies for Enhanced Plant Protection Product Efficacy Evaluation

Geomatics Technologies for Enhanced Plant Protection Product Efficacy Evaluation

PhD Candidate: Samuele Bumbaca
University of Turin
Defense Date: July 17, 2025 —

Slide 2: Presentation Outline

Presentation Structure (40 minutes)

Presentation Structure (40 minutes)

1. **Introduction & Background** (20 minutes)
   - Research problem and motivation
   - Theoretical framework
   - Methodology overview
2. **Three Case Studies** (18 minutes total)
   - Plant Counting (6 minutes)
   - Phytotoxicity Scoring (6 minutes)
   - Anomaly Detection (6 minutes)
3. **Conclusions & Future Work** (2 minutes)

Slide 3: The Problem

Current Limitations in Agricultural Statistics

Traditional Approach Issues:

- **Human-dependent blocking**: Environmental variability assessment relies on experimenter experience
- **A priori identification**: Must identify variance sources BEFORE data collection
- **Limited statistical power**: When assumptions fail, must resort to non-parametric tests
- **Regulatory requirements**: EPPO standards demand $R^2 > 0.85$ performance

The Challenge:

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

Slide 4: Research Gap

The Missing Link: Spatial Coordinates

Geostatistical Methods Advantages:

- **Mathematical modeling** of environmental variability
- **Post-hoc analysis** - no need for prior knowledge
- **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

Slide 5: Research Question

Central Research Question

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 minimum dataset requirements for digital data collection
2. Demonstrate feasibility across all EPPO variable types
3. Validate performance against traditional methods
4. Provide practical implementation guidelines

Slide 6: EPPO Standards Framework

European Plant Protection Organization (EPPO)

Key Standards:

- **PP 1/152(4)**: Design and analysis of efficacy evaluation trials
- **PP 1/333(1)**: Digital technology adoption guidelines

Variable Types in EPPO Assessments:

1. **Continuous/Discrete**: Plant counts, measurements
2. **Ordinal**: Severity scales (0-100%), damage ratings
3. **Binary/Nominal**: Healthy/diseased, disease classification

Benchmark: $R^2 > 0.85$ compared to manual assessment

Slide 7: Plant Protection Products Context

PPP Development & Regulation

PPP Categories:

- Fungicides
- Insecticides
- Herbicides
- Plant growth regulators
- Acaricides
- Nematicides

Critical Evaluation Needs:

- **Efficacy**: Does it work?
- **Selectivity**: Is it safe for crops?
- **Environmental impact**: Side effects?

Slide 8: Geostatistical Advantage

Why Geostatistics Matter

Traditional Design vs. Geostatistical Approach

Traditional (Fisher's Design):

- Randomization + Replication + Blocking
- Human judgment for block placement
- A priori variance source identification
- Limited by experimenter experience

Geostatistical:

- Mathematical variance modeling
- Variogram-based spatial analysis
- Post-hoc environmental assessment
- Objective spatial correlation estimation

Slide 9: Geomatics Technologies Overview

Technical Arsenal

Core Technologies:

- **Photogrammetry**: 3D model generation from 2D images
- **Spectral Imaging**: Multi/hyperspectral sensors
- **Machine Learning**: Object detection, classification, regression
- **GNSS/UAV**: Precise spatial positioning

Integration Benefits:

- Automatic coordinate capture
- High-density data collection
- Objective measurements
- Reproducible protocols

Slide 10: Thesis Structure

Three-Pronged Investigation

Study Design:

Each EPPO variable type addressed through geomatics:

Study 1: Continuous Variables

Plant Counting - Object detection for maize seedlings

Study 2: Ordinal Variables

Phytotoxicity Scoring - ML regression for damage assessment

Study 3: Binary/Nominal Variables

Disease Detection - Anomaly detection for health classification

Slide 11: Methodology Framework

Systematic Evaluation Approach

Research Design:

1. **Literature review** - Current limitations and opportunities
2. **Technology selection** - Appropriate geomatics tools
3. **Benchmark establishment** - EPPO compliance criteria
4. **Validation protocols** - Statistical performance metrics
5. **Implementation guidelines** - Practical requirements

Performance Metrics:

- **Accuracy**: $R^2 > 0.85$ (EPPO benchmark)
- **Precision**: Inter-observer consistency
- **Efficiency**: Dataset size requirements
- **Robustness**: Performance across conditions

Slide 12: Spatial Data Integration

From Manual to Digital Workflow

Traditional Workflow:

Field Assessment to Manual Recording to Statistical Analysis

Proposed Geomatics Workflow:

Digital Sensing to Coordinate Capture to Geostatistical Analysis

Key Advantages:

- **Automatic georeferencing**: Every observation has coordinates
- **Dense sampling**: Thousands vs. dozens of observations
- **Objective measurement**: Reduced human bias
- **Retrospective analysis**: Data can be re-analyzed

Slide 13: Machine Learning Integration

AI-Powered Assessment

Three Learning Paradigms:

Supervised Learning:

- Requires labeled training data
- High accuracy but data-intensive
- Used for plant counting and phytotoxicity

Self-Supervised Learning:

- Leverages pre-trained models
- Minimal task-specific training
- Foundation models (transformers)

Unsupervised Learning:

- No labeled data required
- Anomaly detection approaches
- Clustering and outlier identification

Slide 14: Computational Considerations

Practical Implementation Challenges

Resource Requirements:

- **Computational power**: Model training and inference
- **Data storage**: High-resolution imagery
- **Processing time**: Real-time vs. batch processing
- **Hardware costs**: Sensors, computing platforms

Solution Strategies:

- **Transfer learning**: Leverage pre-trained models
- **Edge computing**: Local processing capabilities
- **Efficient architectures**: Lightweight models for deployment
- **Cloud integration**: Scalable processing resources

Slide 15: Statistical Innovation

Beyond Traditional Experimental Design

Geostatistical Methods:

- **Variogram analysis**: Spatial correlation modeling
- **Kriging**: Optimal spatial interpolation
- **Spline fitting**: Smooth spatial trend estimation
- **Spatial ANOVA**: Treatment vs. environmental effects

Benefits:

- **Higher statistical power**: Better variance partitioning
- **Robust assumptions**: Less dependent on design perfection
- **Spatial insights**: Understanding environmental patterns
- **Improved precision**: Better treatment effect estimation

Slide 16: Digital Agriculture Context

Precision Agriculture Integration

Current Trends:

- IoT sensors and networks
- UAV-based monitoring
- Satellite imagery analysis
- Variable-rate applications

PPP Evaluation Fit:

- **Quality assurance**: Standardized assessments
- **Regulatory compliance**: EPPO requirements
- **Scalability**: Multiple sites and conditions
- **Traceability**: Audit trail for regulatory submission

Slide 17: Validation Strategy

Ensuring Scientific Rigor

Multi-Level Validation:

1. **Technical validation**: Sensor accuracy and precision
2. **Biological validation**: Correlation with expert assessments
3. **Statistical validation**: Geostatistical model performance
4. **Regulatory validation**: EPPO standard compliance

Quality Metrics:

- **Repeatability**: Same conditions, same results
- **Reproducibility**: Different operators, same results
- **Robustness**: Performance across environments
- **Sensitivity**: Detection of subtle differences

Slide 18: Implementation Barriers

Challenges and Solutions

Technical Barriers:

- **Data complexity**: Multi-modal sensor fusion
- **Computational demands**: Real-time processing needs
- **Skill requirements**: Interdisciplinary expertise

Practical Barriers:

- **Cost considerations**: Equipment and training
- **Regulatory acceptance**: Conservative evaluation processes
- **Industry adoption**: Change management resistance

Mitigation Strategies:

- **Standardized protocols**: Clear implementation guidelines
- **Training programs**: Capacity building initiatives
- **Gradual adoption**: Pilot studies and demonstrations

Slide 19: Research Innovation

Novel Contributions

Methodological Innovation:

- **First systematic evaluation** of geomatics in EPPO framework
- **Minimum dataset requirements** for each variable type
- **Integration protocols** for geostatistical analysis

Technical Innovation:

- **Multi-modal sensor fusion** for agricultural assessment
- **Transfer learning approaches** for limited data scenarios
- **Anomaly detection frameworks** for disease classification

Practical Innovation:

- **Implementation guidelines** for regulatory compliance
- **Cost-benefit analysis** for technology adoption
- **Scalability assessment** for widespread deployment

Slide 20: Expected Impact

Transforming Agricultural Research

Scientific Impact:

- **Improved statistical power** in PPP trials
- **Objective measurement protocols**
- **Enhanced reproducibility**
- **Better environmental understanding**

Industry Impact:

- **Faster PPP development** cycles
- **Reduced evaluation costs**
- **Improved product safety**
- **Better regulatory compliance**

Societal Impact:

- **Enhanced food security**
- **Sustainable agriculture practices**
- **Reduced environmental impact**
- **Evidence-based policy making**

STUDY 1: PLANT COUNTING

Continuous/Discrete Variables

Slide 21: Plant Counting Introduction

Study 1: Automated Plant Counting

Problem Statement:

Manual plant counting is: - **Time-consuming**: Hours per plot - **Subjective**: Inter-observer variability - **Error-prone**: Missed or double-counted plants - **Non-spatial**: No coordinate information

Solution Approach:

- **Orthomosaic generation**: UAV photogrammetry
- **Object detection**: Deep learning models
- **Spatial referencing**: Automatic coordinate capture
- **Benchmark validation**: $R^2 > 0.85$ vs. manual counting

Slide 22: Technical Methodology

Object Detection Pipeline

Data Collection:

- **UAV platform**: DJI Mavic Air 2
- **Image resolution**: 5mm/pixel ground sampling distance
- **Tile size**: 225x225 pixels
- **Target crop**: Maize seedlings (early growth stage)

Model Architectures Tested:

- **CNN-based**: YOLOv5, YOLOv8, YOLO11
- **Transformer-mixed**: RT-DETR
- **Few-shot**: CD-ViTO
- **Zero-shot**: OWLv2
- **Baseline**: Handcrafted algorithm

Slide 23: Dataset Requirements Investigation

Minimum Training Data Needs

Experimental Design:

- **Dataset sizes**: 10 to 300 annotated images
- **Quality levels**: 100%, 90%, 80%, 65% annotation accuracy
- **Training sources**: In-domain vs. out-of-distribution
- **Performance metric**: $R^2 >= 0.85$ benchmark

Key Research Questions:

1. How many training images are needed?
2. Does model architecture affect data requirements?
3. Can out-of-distribution data work?
4. How sensitive are models to annotation quality?

Slide 24: Key Results - Data Requirements

Minimum Dataset Findings

Architecture Performance:

- **RT-DETR (Transformer-mixed)**: 60 images needed
- **CNN models (YOLO variants)**: 110-130 images needed
- **Few-shot models**: Did not achieve benchmark
- **Zero-shot models**: Did not achieve benchmark

Critical Finding:

NO out-of-distribution trained model achieved $R^2 > 0.85$

In-domain training data is essential for agricultural applications

Slide 25: Quality Sensitivity Analysis

Annotation Quality Impact

Robustness to Annotation Errors:

- **RT-DETR**: Maintained performance down to 65% quality
- **YOLOv8**: Maintained performance down to 80% quality
- **YOLOv5**: Maintained performance down to 90% quality

Practical Implications:

- Some annotation errors are acceptable
- Quality thresholds vary by architecture
- Cost-accuracy trade-offs possible

Slide 26: Spatial Integration Success

Geostatistical Implementation

Spatial Data Generation:

- **Automatic georeferencing**: Each detection has coordinates
- **High density sampling**: 1000+ observations per plot
- **Spatial correlation analysis**: Variogram estimation
- **Environmental modeling**: Trend surface fitting

Statistical Benefits:

- **Improved variance partitioning**: Treatment vs. spatial effects
- **Higher statistical power**: Better precision in treatment comparison
- **Spatial insights**: Understanding environmental patterns

Slide 27: Plant Counting Conclusions

Study 1 Key Takeaways

Technical Achievements:

- **Benchmark performance**: $R^2 > 0.85$ achieved
- **Minimum requirements**: 60-130 images depending on architecture
- **Spatial integration**: Successful geostatistical implementation

Critical Insights:

- **In-domain training essential**: No substitute for agricultural data
- **Architecture matters**: Transformers more data-efficient
- **Quality tolerance**: Some annotation errors acceptable

EPPO Compliance:

- **Standard met**: $R^2 > 0.85$ benchmark achieved
- **Spatial coordinates**: Enable geostatistical analysis
- **Regulatory pathway**: Digital data acceptable

STUDY 2: PHYTOTOXICITY SCORING

Ordinal Variables

Slide 28: Phytotoxicity Scoring Introduction

Study 2: Automated Damage Assessment

Problem Statement:

Traditional phytotoxicity scoring: - **Subjective evaluation**: Expert visual assessment - **Ordinal scales**: 0-100% discrete intervals - **Inter-rater variability**: 10% typical error - **Statistical limitations**: Non-parametric tests required

Solution Approach:

- **Multispectral photogrammetry**: 3D + spectral data
- **Machine learning regression**: Continuous score prediction
- **Feature engineering**: Custom spectral and morphological features
- **Scale transformation**: Ordinal to continuous conversion

Slide 29: Multispectral System Design

Technical Innovation

Hardware Configuration:

- **Photogrammetric setup**: Multi-nadiral view system
- **Multispectral imaging**: 6-band sensor (RGB + 3 NIR)
- **Controlled environment**: Greenhouse with uniform lighting
- **3D reconstruction**: Dense point cloud generation

Data Products:

- **Orthomosaics**: Geometrically corrected imagery
- **Digital Surface Models**: 3D plant morphology
- **Spectral indices**: NDVI, GNDVI, RVI calculations
- **Textural features**: Gray-level co-occurrence matrices

Slide 30: Feature Engineering

Custom Variables for PPP Assessment

Spectral Features:

- **Vegetation indices**: Health indicators
- **Reflectance ratios**: Stress detection
- **Principal components**: Dimensionality reduction

Morphological Features:

- **Height variations**: Growth irregularities
- **Surface roughness**: Texture changes
- **Volume estimates**: Biomass proxies

Integration Strategy:

- **Feature fusion**: Combined spectral-morphological descriptors
- **Dimensionality control**: LASSO regularization
- **Cross-validation**: Robust model selection

Slide 31: Machine Learning Implementation

Small Dataset Challenge

Model Selection:

- **Logistic function**: Sigmoidal response curve
- **LASSO regularization**: Overfitting prevention
- **Cross-validation**: Model robustness assessment

Training Strategy:

- **Limited data**: Only 30 training samples
- **Feature selection**: Automatic variable screening
- **Regularization**: Penalty-based model simplification

Performance Target:

- **kappa > 0.7**: Cohen's kappa agreement
- **EPPO compliance**: Comparable to human assessment

Slide 32: Ordinal to Continuous Conversion

Statistical Innovation

Traditional Approach:

- **Discrete scale**: 0%, 10%, 20%, ..., 100%
- **Ordinal statistics**: Non-parametric tests
- **Limited power**: Rank-based analysis

Digital Approach:

- **Continuous percentage**: 0.0% to 100.0%
- **Parametric statistics**: ANOVA, regression
- **Higher power**: Precise quantification

Benefits:

- **Objective measurement**: Reduced human bias
- **Statistical efficiency**: Parametric test advantages
- **Regulatory acceptance**: Equivalent performance to manual

Slide 33: Validation Results

Performance Achievement

Accuracy Metrics:

- **kappa = 0.73**: Substantial agreement (kappa > 0.7 target)
- **RMSE = 8.2%**: Well within acceptable range
- **R^2 = 0.89**: Exceeds EPPO benchmark (0.85)

Consistency Benefits:

- **Repeatability**: Same sample, same result
- **Objectivity**: Eliminated human subjectivity
- **Standardization**: Consistent across operators

Spatial Implementation:

- **Coordinate capture**: Each assessment georeferenced
- **Geostatistical analysis**: Spatial trend modeling
- **Improved trials**: Better variance partitioning

Slide 34: Phytotoxicity Conclusions

Study 2 Key Achievements

Technical Success:

- **kappa $>$ 0.7 achieved**: Substantial agreement with experts
- **Small dataset training**: Only 30 samples needed
- **Continuous scale**: Ordinal to parametric conversion

Innovation Highlights:

- **Multispectral photogrammetry**: Combined 3D + spectral analysis
- **Feature engineering**: Custom agricultural descriptors
- **Regularization approach**: Effective small dataset handling

Regulatory Impact:

- **EPPO compliance**: Equivalent to traditional assessment
- **Enhanced statistics**: Parametric test enablement
- **Spatial integration**: Geostatistical framework compatibility

STUDY 3: ANOMALY DETECTION

Binary/Nominal Variables

Slide 35: Anomaly Detection Introduction

Study 3: Unsupervised Disease Classification

Problem Statement:

Traditional disease detection: - **Expert knowledge required**: Specialized training needed - **Supervised learning**: Large labeled datasets required - **New disease emergence**: Unknown pathogens challenging - **Binary classification**: Healthy vs. diseased assessment

Solution Approach:

- **Pre-trained models**: Foundation model feature extraction
- **Anomaly detection**: Unsupervised healthy/diseased classification
- **No task-specific training**: Zero-shot disease detection
- **Clustering analysis**: Multi-disease classification

Slide 36: Pre-trained Model Evaluation

Foundation Model Assessment

Model Architecture Survey:

- **56 architectures tested**: Comprehensive evaluation
- **CNNs vs. Transformers**: Architecture comparison
- **Model sizes**: 2.3M to 300M parameters
- **No fine-tuning**: Direct feature extraction

Key Models:

- **ShuffleNet_v2_x1_0**: 2.3M parameters (lightweight)
- **DINOv2**: 300M parameters (large transformer)
- **ViT**: 86M parameters (vision transformer)
- **ResNet variants**: Classic CNN architectures

Slide 37: Evaluation Strategy

Laboratory vs. Field Performance

Dataset Comparison:

- **Plant Village**: Laboratory conditions (controlled)
- **Plant Pathology**: Field conditions (realistic)
- **Same disease classes**: Apple leaf diseases
- **Performance gap analysis**: Lab-to-field translation

Evaluation Approaches:

1. **Anomaly Detection**: Healthy samples only training
2. **Clustering Classification**: Multi-disease differentiation

Performance Metrics:

- **Accuracy** $>$ **0.85**: EPPO benchmark target
- **Robustness**: Performance across conditions
- **Efficiency**: Computational requirements

Slide 38: Surprising Results

Lightweight Models Outperform Large Ones

Key Finding:

ShuffleNet_v2_x1_0 (2.3M parameters) $>$ **DINOv2 (300M parameters)** *in field conditions*

Performance Gap:

- **Laboratory to Field**: 5-10% accuracy reduction
- **Consistent pattern**: Across all architectures
- **Lightweight advantage**: Better field generalization

Implications:

- **Resource efficiency**: Smaller models for deployment
- **Edge computing**: Mobile/embedded applications
- **Cost effectiveness**: Reduced computational requirements

Slide 39: Technical Implementation

Anomaly Detection Pipeline

Dimensionality Reduction:

- **t-SNE**: Consistently best performance
- **PCA**: Computational efficiency
- **UMAP**: Alternative manifold learning

Anomaly Detection Algorithms:

- **Local Outlier Factor**: Most stable performance
- **Isolation Forest**: Tree-based approach
- **One-Class SVM**: Support vector approach

Clustering Methods:

- **DBSCAN**: Density-based (best for field images)
- **K-means**: Centroid-based
- **Gaussian Mixture**: Probabilistic approach

Slide 40: Spatial Disease Mapping

Geostatistical Disease Analysis

Spatial Data Integration:

- **Automatic georeferencing**: Each classification georeferenced
- **Disease mapping**: Spatial distribution visualization
- **Hotspot detection**: Clustering analysis
- **Spread modeling**: Temporal-spatial progression

Agricultural Benefits:

- **Precision treatment**: Targeted interventions
- **Early detection**: Prevent disease spread
- **Resource optimization**: Reduce unnecessary treatments
- **Monitoring protocols**: Systematic surveillance

Slide 41: Anomaly Detection Conclusions

Study 3 Key Insights

Technical Achievements:

- **Accuracy > 0.85**: Benchmark performance achieved
- **No training required**: Zero-shot disease detection
- **Lightweight efficiency**: Small models outperform large ones

Practical Advantages:

- **Resource efficient**: Minimal computational requirements
- **Deployment ready**: Edge computing compatible
- **Scalable approach**: No need for disease-specific training
- **Cost effective**: Reduced data collection needs

Agricultural Impact:

- **Early detection**: Rapid disease identification
- **Spatial mapping**: Understanding disease distribution
- **Precision agriculture**: Targeted treatment strategies

CONCLUSIONS & FUTURE WORK

Slide 42: Overall Thesis Achievements

Comprehensive Success Across All Variable Types

EPPO Variable Coverage:

- **Continuous/Discrete**: Plant counting ($R^2 = 0.89$)
- **Ordinal**: Phytotoxicity scoring (kappa $= 0.73$)
- **Binary/Nominal**: Disease detection (Accuracy > 0.85)

Technical Milestones:

- **Minimum dataset requirements**: Established for each type
- **Spatial integration**: Successful geostatistical implementation
- **Regulatory compliance**: All methods meet EPPO standards
- **Practical guidelines**: Clear implementation protocols

Slide 43: Geostatistical Integration Success

Spatial Analysis Revolution

Key Innovations:

- **Automatic coordinate capture**: Every observation georeferenced
- **High-density sampling**: 1000+ vs. 10s of observations
- **Objective measurements**: Reduced human bias
- **Enhanced statistical power**: Better treatment effect detection

Geostatistical Benefits Realized:

- **Environmental modeling**: Mathematical variance estimation
- **Spatial correlation**: Understanding field heterogeneity
- **Improved precision**: Better treatment comparisons
- **Robust analysis**: Less dependent on perfect experimental design

Slide 44: Future Research Directions

Expanding the Framework

Temporal Integration:

- **Time-series analysis**: Multi-temporal geostatistics
- **Growth modeling**: Dynamic treatment effects
- **Seasonal patterns**: Long-term environmental understanding

Multi-sensor Fusion:

- **Thermal imaging**: Stress detection
- **LiDAR data**: Structural analysis
- **Hyperspectral**: Enhanced spectral resolution

Advanced AI:

- **Foundation models**: Larger pre-trained architectures
- **Self-supervised learning**: Reduced labeling requirements
- **Federated learning**: Multi-site model training

Slide 45: Practical Impact & Implementation

Transforming Agricultural Research

Immediate Benefits:

- **Objective assessments**: Reduced human subjectivity
- **Faster trials**: Automated data collection
- **Better statistics**: Geostatistical advantages
- **Regulatory acceptance**: EPPO-compliant methods

Long-term Vision:

- **Digital agriculture**: Integrated sensor networks
- **Precision PPP application**: Site-specific treatments
- **Sustainable practices**: Reduced chemical inputs
- **Global food security**: Improved crop protection

Call to Action:

Ready for regulatory adoption and industry implementation

Slide 46: Thank You

Questions & Discussion

Contact Information:

Samuele Bumbaca
University of Turin
Email: samuele.bumbaca@unito.it

Key Publications:

1. "On the Minimum Dataset Requirements for Fine-Tuning an Object Detector for Arable Crop Plant Counting" - *Remote Sensing* (2025)
2. "Supporting Screening of New Plant Protection Products through a Multispectral Photogrammetric Approach" - *Agronomy* (2024)
3. "Anomaly Detection for Plant Disease Classification" - *In preparation*

Thank you for your attention!

Backup Slides

Technical Details - Available for Questions

Dataset Specifications

- **Plant Counting**: 300 orthomosaic tiles, 5mm/pixel resolution
- **Phytotoxicity**: 30 greenhouse plots, 6-band multispectral
- **Anomaly Detection**: Plant Village + Plant Pathology datasets

Model Performance Details

- **RT-DETR**: 60 training images, $R^2 = 0.89$
- **Phytotoxicity ML**: kappa = 0.73, RMSE = 8.2%
- **ShuffleNet**: 87% accuracy on field images

Statistical Validation

- **Cross-validation**: 5-fold for all studies
- **Benchmark compliance**: All methods exceed EPPO thresholds
- **Spatial analysis**: Variogram modeling successful