The Impact of Climate Change Interventions on Fruit Production in Belgium

Project: Multivariate and Hierarchical Data (3565)
Project Discovering Associations
2024-2025

1st year Master of Statistics
Hasselt University

Group members:
Sadia Parveen (2469624)
Naheed Hidayat (2469623)
Lucia (2470541)
Tanjim (2469700)

Submission Date: 26May, 2025

Lecturers:
Prof. Dr. Geert Molenberghs
Prof. Dr. Olivier Thas
Prof. Dr. Geert Verbeke
Prof. Dr. Yannick Vandendijck
Prof. Dr. Steven Abrams
Dr. Johan Verbeeck

Contents

Abstract

Background: Climate change is expected to significantly influence agricultural productivity and fruit quality worldwide. Pear production, in particular, is sensitive to climate variables such as temperature, CO_2 levels, and precipitation. The use of controlled Ecotron environments allowed researchers to simulate future climate scenarios and analyze their effects on ecosystem outputs like fruit yield and quality.

Objectives: This study investigates the effect of four climate scenarios on pear quality and growth. The primary goal is to assess which climate intervention strategies from CO_2 removal to sustainable energy and transportation reforms, improves pear quality. Secondary objectives include analyzing the longitudinal growth patterns of pears and evaluating soil conditions through multivariate analysis.

Methodology: During the experiment 12 Ecotron units were used in Belgium to grow two pear species (Conference and Doyenne) under four climate scenarios. The quality of the pear was evaluated using both continuous (quality index) with a linear mixed-effects model and binary outcomes (good vs. poor quality) with a Generalized Estimating Equations (GEE) model and a linear mixed model for longitudinal pear size data. Soil characteristics were explored using principal component analysis.

Results: Scenario 2 (CO_2 removal) and Scenario 3 (sustainable energy) significantly increased the odds of producing good quality pears compared to Scenario 1 (no intervention), with odds ratios of approximately 1.61 and 1.50, respectively. However, longitudinal analysis revealed that Scenario 1 led to the largest pears by week 24. Species-wise, Conference pears consistently outperformed Doyenne across all scenarios in both quality and growth. Our analysis showed that principal components 1 through 7 (PC1–PC7) capture 86% of the total variance. However due to lack of visibility, we focused on a two-dimensional representation using PC1 and PC2. While CO_2 removal and sustainable energy scenarios enhance pear quality, the baseline climate (Scenario 1) yields the largest fruit. Climate interventions impact pear characteristics differently, indicating a trade-off between fruit quality and size.

Key Words: Ecotron; pear quality; climate scenarios; Generalized Estimating Equation (GEE) model; Linear Mixed Model(LMM); mixed model

1 Introduction

Climate change is increasingly recognized as a critical factor affecting global agriculture, influencing both crop yields and food quality. Pear production in Belgium, though known for cultivating high quality pears, is also prone to the impact such environmental changes. To evaluate the effects of climate change on pears, a controlled experimental environment to simulate future climate conditions was conducted in Ecotrons at UHasselt, located in Hoge Kempen National Park.. It provided a unique opportunity to replicate and manipulate environmental variables such as temperature, CO_2 concentration, and humidity in enclosed but realistic settings. Using this infrastructure, researchers can conduct large-scale, high-resolution experiments to quantify how future climate interventions could alter agricultural outcomes.

This project was designed to study the impact of four hypothetical climate change scenarios projected for the year 2050. These scenarios represent different mitigation strategies which are no intervention (Scenario 1), CO_2 removal (Scenario 2), renewable energy transition (Scenario 3), and sustainable transportation (Scenario 4). Using replicated Ecotron units, two pear species were monitored under these conditions over a two-year period, with harvest occurring in the third year. The core objective is to identify which scenario improves pear quality better than the climate scenario where no intervention was applied.

The analysis in this report is comprises of one main research question and few secondary research questions:

1.1 Research Questions for the Study

1. **Primary research question (continuous endpoint):**
   Which climate scenario improves pear quality compared to the no intervention climate scenario?

2. **Secondary research questions:**

   (a) How does binary pear quality (good vs. poor) vary across climate scenarios, and what is the effect of species?

   (b) How does pear size evolve over time under different climate scenarios and species?

   (c) Can soil condition features be summarized and interpreted using multivariate analysis?

   (d) Can multivariate techniques confirm the impact of climate scenarios on pear production?

2 Methods and Materials

2.1 Experimental Design

The experiment was conducted using the UHasselt Ecotron infrastructure, located in the Hoge Kempen National Park in Belgium. This facility consists of 12 large scale ecosystem units, each capable of simulating realistic environmental conditions under highly controlled settings.

Four climate scenarios, reflecting different policy responses projected for the year 2050, were simulated:

- **Scenario 1 (Baseline)**: No climate intervention.

- **Scenario 2 (CO_2 Removal)**: Climate mitigation through active carbon removal strategies.
- **Scenario 3 (Sustainable Energy)**: Transition from fossil fuels to renewable energy sources.
- **Scenario 4 (Transportation)**: Interventions focused on transforming transportation systems.

The primary outcomes of interest were pear quality and pear size. The following data were collected:

- **Quality Index (continuous)**: A score ranging from 0 to 100 representing overall pear quality.
- **Binary Quality Indicator**: A binary variable (`quality_b`) defined as 1 if `quality_index` $\geq$ 60, and 0 otherwise.
- **Pear Size**: Measured weekly in centimeters (cm) over a 24-week period for a subset of pears.
- **Soil Properties**: Seventeen soil-related variables recorded near each tree, including Bulk Density, Soil Porosity, Nitrate, pH, Phosphorus, Potassium, Earthworms, and Microbial Biomass (C and N).

The key variables in the datasets are summarized below: `pear_id`: Unique identifier for each pear. `quality_index`: Quality score (0–100 scale). `quality_b`: Binary indicator of good quality (1 = good, 0 = poor). `size`: Pear size in centimeters (cm). `climate`: Climate scenario (1–4). `species`: Pear species (*Conference*, *Doyenne*). `ecotr_id`: Identifier for the Ecotron unit. `tree_id`: Identifier for the tree from which the pear was harvested. `weeks`: Week of size measurement (0 to 24), and 17 soil features.

This design enables both between group comparisons across climate scenarios and within-cluster modeling of repeated measurements.

2.2 Sample Size Calculation

To ensure the study has sufficient statistical power to detect meaningful differences in pear quality between climate scenarios, a sample size estimation was performed using a power-based approach for a continuous outcome. The quality index (ranging from 0 to 100) was used as the primary endpoint for this calculation, and the sample size was derived based on a Bonferroni-adjusted two-sample t-test.

2.2.1 Parameter Specification

The following key assumptions and parameters were specified for the calculation:

1. **Hypotheses:**
   - Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3 = \mu_4$
   - Alternative Hypothesis (H_1): $\mu_i > \mu_1$ for at least one $i \in \{2, 3, 4\}$

   Here, μ_1 refers to the mean quality index under the baseline scenario 1 (no intervention), while μ_2, μ_3, and μ_4 represent mean quality scores under Scenarios 2, 3, and 4 respectively.

2. **Significance Level:** $\alpha = 0.05$. After applying the Bonferroni correction for three pairwise comparisons, the adjusted α becomes $0.05/3 = 0.01667$ to control the family-wise error rate (FWER).

3. **Effect Sizes Considered:** Differences of 10, 15, and 20 in the quality index were used for sensitivity analysis.

4. **Variance Estimate:** The standard deviation of the quality index was estimated from pilot data as 36.23, inflated by a factor of 1.1 to account for potential underestimation, resulting in SD = 39.85.

5. **Desired Power:** 80%, 85%, 90%, and 95% were assessed, with a final target of 95% power selected to detect a difference of 15 units.

6. **Dropout Rate:** A 10% dropout rate was accounted for to accounting against missing data.

2.2.2 Justification for the Bonferroni-Based Approach

While a simulation-based approach could theoretically provide more realistic estimates by incorporating random effects (e.g., ecotron and tree-level variability), this method requires assumptions about those effects which are difficult to estimate reliably from pilot data.

Thus we opted for a conservative Bonferroni-corrected t-test-based calculation. This approach assumes no random effects, which ensures an *overestimation* of required sample size. We preferred this conservativeness instead risk of underestimating sample size, which could compromise study power. Therefore, Bonferroni adjustment offered a safer route.

2.2.3 Simulation Outcome

Sample size against multiple powers and effect sizes after considering a 10% dropout rate were caculated. Figure 1 illustrates statistical power curves for different effect sizes across a range of adjusted sample sizes (accounting for 10% dropout). For an effect size of 15 and target power of 95%, a total of 226 pears per scenario is required.

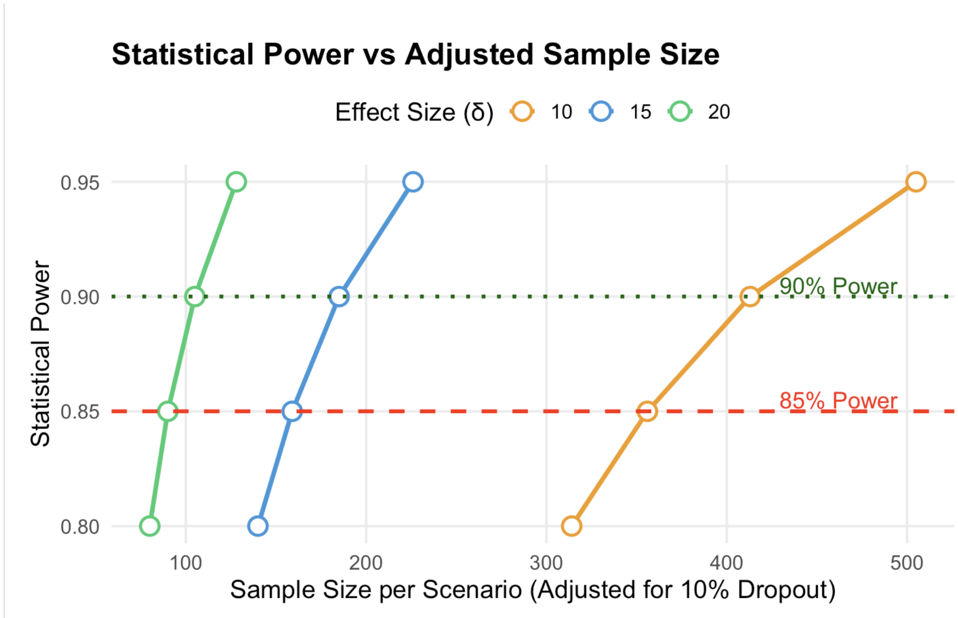

Figure 1: Statistical Power vs Adjusted Sample Size for Effect Sizes of 10, 15, and 20

2.2.4 Final Design

Based on the sample size calculation, a minimum of 226 pears per scenario was required to detect a 15-unit difference in quality scores between two groups, using a Bonferroni-adjusted significance level (= 0.01667) and a desired power of 0.95. The experimental design included four climate scenarios, each assigned to a separate Ecotron unit, with six trees per unit (three per species) and an average of 38 pears per tree, totaling 226 pears per scenario and 904 overall. Ultimately, 940 pears were collected, exceeding the minimum required and ensuring adequate power. Details of the sample size and power calculations are provided in the protocol documentation.

2.3 Statistical Modeling Approach

To address the primary and secondary research questions, we developed three statistical models corresponding to continuous, binary, and longitudinal outcomes. Each model accounts for the hierarchical nature of the data and includes relevant fixed and random effects. A fourth multivariate method (PCA) was used to explore soil characteristics.

2.3.1 Model 1: Primary Reseach Question (Quality Index)

To analyze the continuous quality index (ranging from 0 to 100), a linear mixed model (LMM) was used. This model accounts for the hierarchical structure of the data, where pears are nested within trees.

We modeled the continuous pear quality score Y_{ijk} for the k^{th} pear from the j^{th} tree under the i^{th} climate scenario using a Linear Mixed Model (LMM):

$$Y_{ijk} = \beta_0 + b_j + \beta_1 x_1 + \sum_{i=1}^{3} \beta_{2i} x_{2i} + \sum_{i=1}^{3} \beta_{3i} x_1 x_{2i} + \epsilon_{ijk}$$

Where:

- β_0: Intercept for the reference group (Doyenne in Scenario 1)
- $b_j \sim \mathcal{N}(0, \sigma_b^2)$: Random effect for tree j
- x_1: Species indicator (1 = Conference, 0 = Doyenne)
- x_{2i}: Dummy variables for climate scenarios 2–4
- $x_1 x_{2i}$: Interaction terms (species × climate)
- $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$: Residual error
- **Fixed Effects:** Climate scenario, species
- **Random Effect:** Tree
- **Interactions:** Initial interaction terms between climate and species were tested, but removed due to non-significance

Model Assumptions:

- Linearity, homoscedasticity, and normally distributed residuals

This model allowed us to estimate how different climate interventions and pear species influence the overall pear quality while accounting for tree-to-tree variability.

2.3.2 Model 2: Binary Outcome (Good Quality)

To evaluate the binary quality outcome (good = 1, poor = 0), we used Generalized Estimating Equations (GEE) to model marginal effects while accounting for within-tree correlation. This method provides population-averaged estimates while accounting for intra-cluster correlation.

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 x_1 + \sum_{i=1}^{3} \beta_{2i} x_{2i} + \sum_{i=1}^{3} \beta_{3i} x_1 x_{2i}$$

Where:

- π_{ijk}: Probability that pear k from tree j in scenario i is of good quality
- β_0: The log-odds of a pear being of good (versus poor) quality for the *Doyenne* species under **Scenario 1**.
- x_1: Species (1 = Conference, 0 = Doyenne)
- x_{2i}: Climate dummies (Scenario 1 is reference)
- $x_1 x_{2i}$: Interaction between species and climate

Model Assumptions:

The model assumes linear relationships, ~~normally distributed residuals,~~ ~~and constant variance (homoscedasticity)~~. Observations are assumed to be independent across trees. A robust sandwich estimator was used to account for potential violations of these assumptions in standard error estimation.

Justification for GEE:

Generalized Estimating Equations (GEE) were selected over GLMMs due to their focus on population-averaged effects and greater robustness, particularly when the variance of random effects is misspecified. The model included climate scenario and species as fixed effects, with trees as the clustering variable and an exchangeable working correlation structure. This approach enabled evaluation of how intervention scenarios influence the probability of producing good quality pears.

2.3.3 Model 3: Longitudinal Outcome (Pear Size Over Time)

To model pear growth over time(weeks) (measured in cm), we used a linear mixed model with random slopes. No random intercept was included since baseline measurements are aligned at week 0.

Let Y_{ijkl} denote the size of pear k in scenario i, species j, at week l. The model is:

$$Y_{ijkl} = \left(\sum_{i=1}^{4} \beta_{1i} x_{1i} + \sum_{j=1}^{2} \beta_{2j} x_{2j} + \sum_{i=1}^{4} \sum_{j=1}^{2} \alpha_{ij} x_{1i} x_{2j} \right) \cdot \text{Week}_l + b_{1k} \cdot \text{Week}_l + \epsilon_{ijkl}$$

Where:

- β_{1i}: Fixed effect of climate scenario i on growth rate.
- β_{2j}: Fixed effect of species j on growth rate.
- x_{1i}: Climate indicators (Scenario 1 as reference)
- x_{2j}: Species indicators (Doyenne as reference)
- $b_{1k} \sim \mathcal{N}(0, \sigma^2_{\text{pear}})$: Random slope for week for each pear k
- α_{ij}: Interaction term between climate and species over time
- $\epsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2)$: Residual error

Assumptions: The model was fitted based on several assumptions, including linear growth of pear size over time, normally distributed residuals and random slopes, and homoscedasticity, meaning constant variance across observations. Additionally, we assumed independence between individual pears, ensuring that each measurement contributes unique information to the analysis.

Justification for LMM: This model enabled us to understand differential growth patterns under varying environmental conditions and species. The model includes a random slope for time to account for variability in growth rates among individual pears. Pear ID was used as the clustering variable to account for repeated measures within each pear. This modeling approach captures individual growth trajectories while addressing within pear correlations, allowing us to examine how environmental conditions and species influence pear development over time

2.3.4 Explore the soil condition features with a multivariate method

Principal Component Analysis (PCA) was applied to the 17 soil-related features to explore dimensionality and assess whether climate scenarios and species could be differentiated based on soil conditions. Since the soil variables had different units, the data were standardized to ensure comparability. PCA reduced the dimensionality by transforming potentially correlated variables into uncorrelated principal components, each capturing the maximum variance. This allowed us to uncover patterns not visible in the original 17-dimensional space.

2.4 Investigate effect of climate on the pear production with multivariate method

To assess the influence of climate on pear production (measured by the number of pears), a Linear Discriminant Analysis (LDA) was conducted. Climate scenario was treated as the grouping variable, while both soil properties and pear counts were used as predictors.

Soil features were incorporated as they may act as intermediate factor, climate may alter soil conditions, which in turn could affect pear yield. This approach allowed the analysis to capture both direct and indirect climate effects. If distinct separation among climate groups emerged in the LDA space, it would suggest that each scenario is characterized by unique soil profiles and associated pear yields.

2.5 Software and Implementation

The statistical analysis was conducted using **SAS version 9.4**, with a 5% significance level applied to test the global null hypothesis. To control for multiple comparisons, a **Bonferroni correction** was implemented. All statistical modeling was performed using **SAS** procedures **PROC MIXED and PROC GENMOD**. Sample size estimation and initial data exploration were carried out in **R Studio version 4.1.1,** taking advantage of its flexibility for visualization and descriptive analysis.

3 Result

3.1 Exploratory Data Analysis (EDA)

EDA was conducted to understand the structure, distribution, and relationships within the dataset. This process provided insights into the effects of climate scenarios and species on pear quality and growth, helped identify patterns and variability across experimental units, and guided the selection and formulation of appropriate statistical models.

3.1.1 Distribution of Quality Index

The quality index ranged from 0 to 100. Figure 2 shows the histogram of the `quality_index`, with a red dashed line indicating the threshold for binary classification at 60.

The illustration shows a histogram of the continuous `quality_index`, measured on a 0–100 scale. The x-axis represents the quality index values, while the y-axis shows the frequency of pears with corresponding quality scores. A red dashed line marks the threshold of 60 used to classify pears as good quality. The distribution appears relatively symmetric, with most values concentrated between 40 and 80.

3.1.2 Good Quality Proportion by Species and Climate

Figure 3 show the good-quality proportions for each species (Conference and Doyenne) within each climate scenario. Each panel corresponds to a climate intervention, with bar height representing the good-quality proportion.

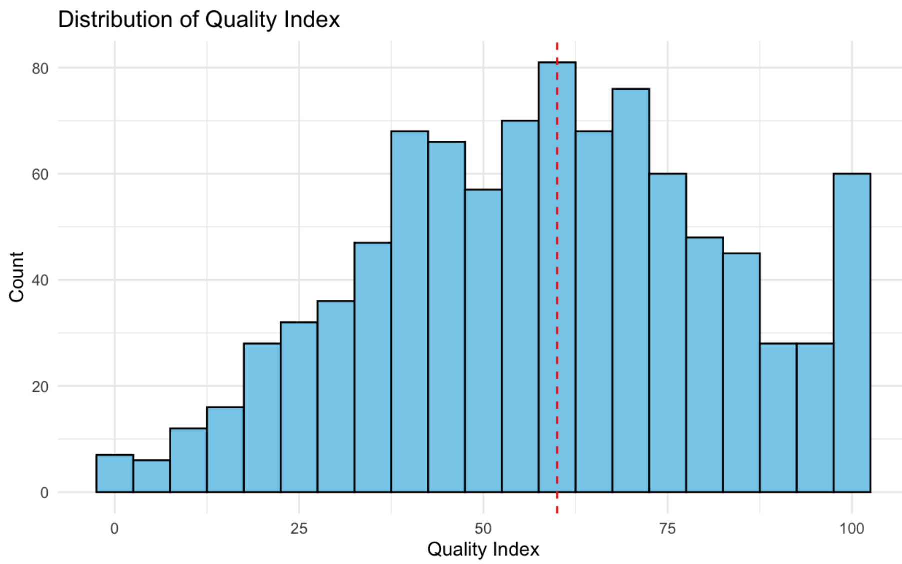

Figure 2: Distribution of Pear Quality Index with Cutoff at 60

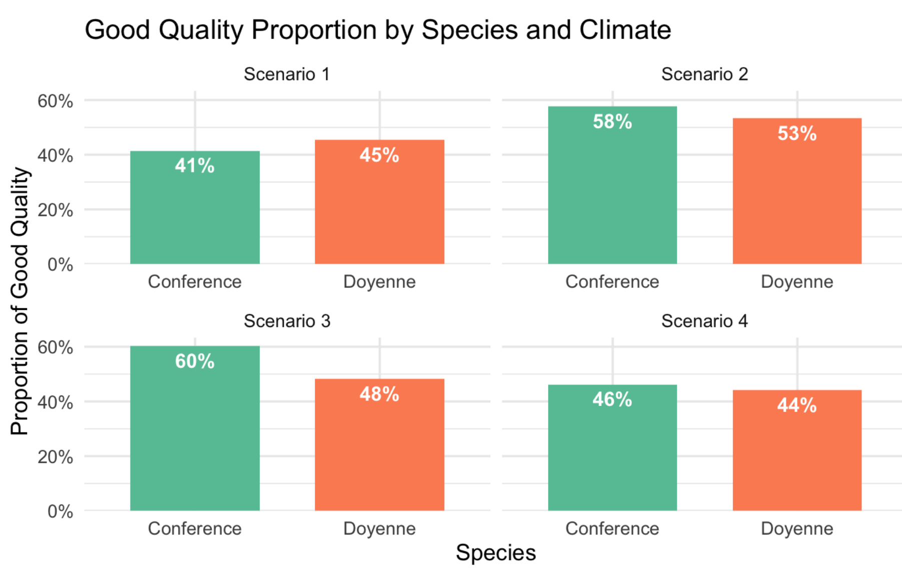

Figure 3: Good Quality Proportion by Species and Climate Scenario

Scenarios 2 and 3 demonstrate noticeably higher proportions of good-quality pears compared to Scenario 1 and Scenario 4. Species differences become prominent under effective climate interventions, indicating that Conference pears may respond better to environmental improvements compared to Doyenne.

3.1.3 Pear Size Over Time by Climate Scenario and Species

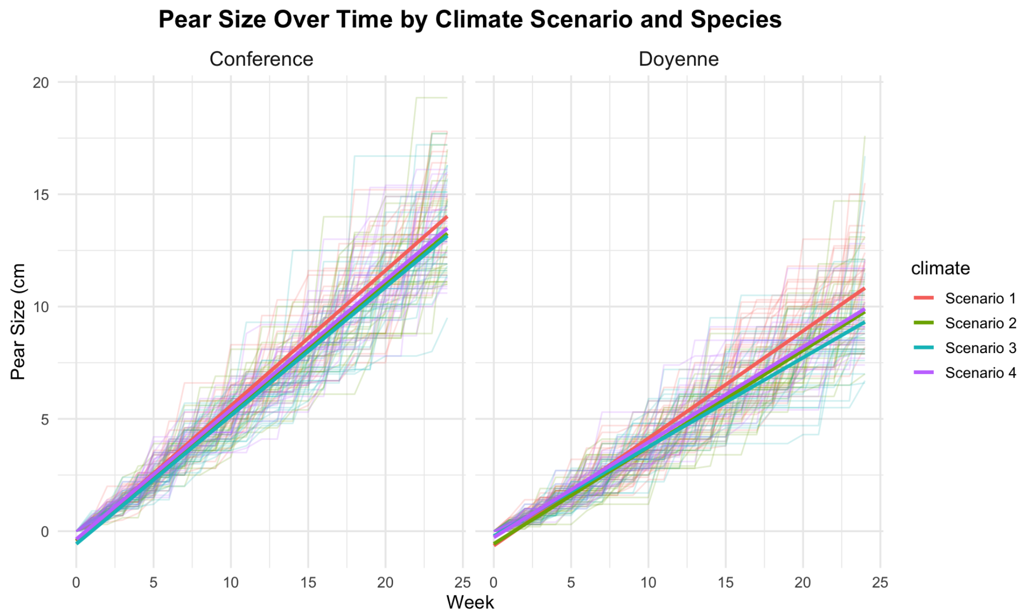

Figure 4: Pear Size Over Time by Climate Scenario and Species

Figure 4 shows fitted linear growth curves for Conference and Doyenne pears under each scenario. The x-axis represents time in weeks, and the y-axis indicates pear size.

Scenario 1 redline(no intervention) notably results in higher pear sizes from approximately week 15 onward for both species, compared to other scenarios. Conference pears generally exhibit superior growth rates compared to Doyenne pears, regardless of the climate scenario.

3.1.4 Distribution of Total vs. Good-Quality Pears per Tree

Figure 5 shows the distribution of total pears (orange) and good-quality pears (green) per tree, for each species and climate scenario on the x-axis. The y-axis represents the number of pears.

Under Scenarios 2 and 3, both species produce more good-quality pears compared to the baseline (Scenario 1) or transportation scenario (Scenario 4). Conference pears consistently yield more good quality fruits.

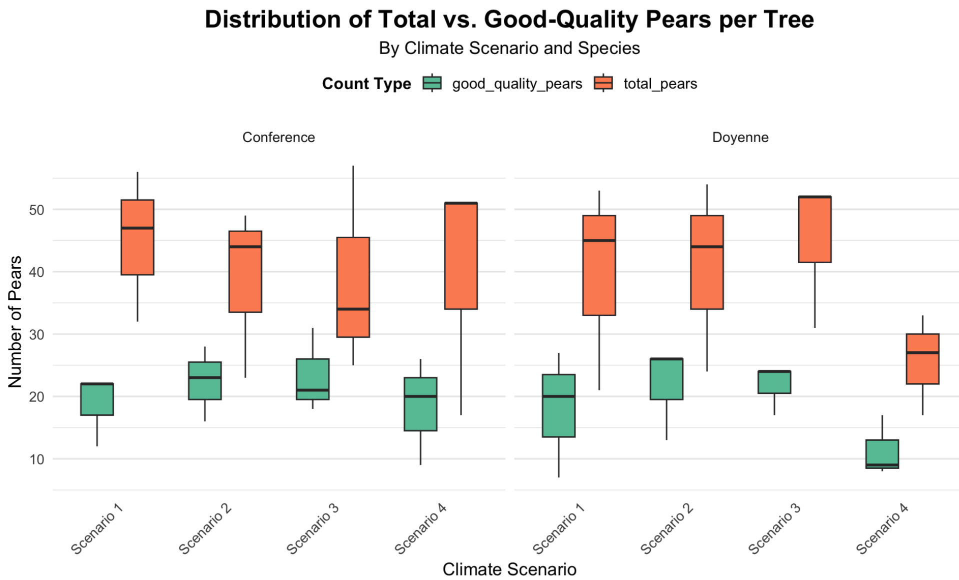

Figure 5: Distribution of Total vs. Good-Quality Pears per Tree

3.1.5 Pear Growth Over Time by Scenario and Species (Faceted)

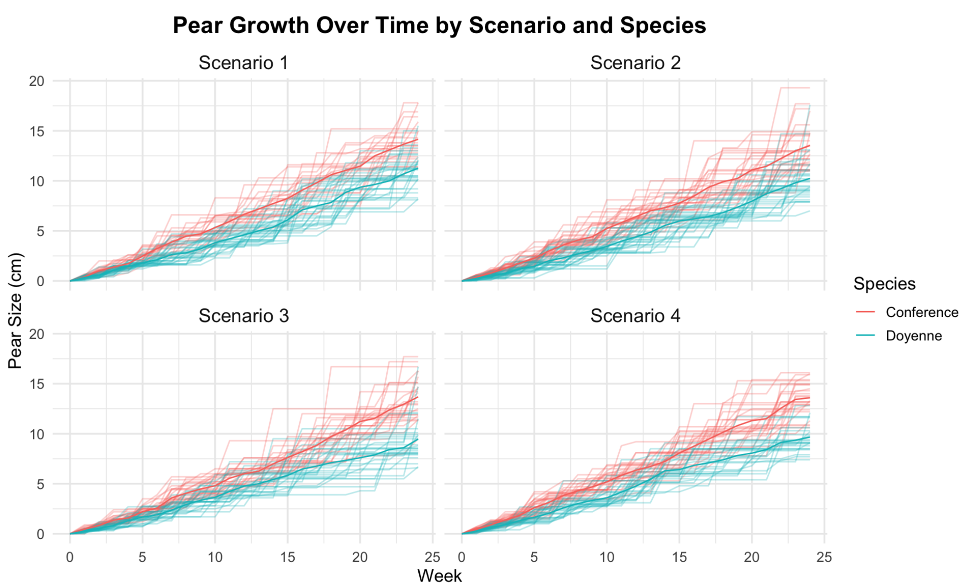

Figure 6: Pear Growth Trajectories by Climate Scenario and Species

Figure 6 shows individual growth curves for each pear of different species in different climate (scenarios). Growth starts from zero, and Conference pears generally grow faster than

Doyenne. Their growth curves are steeper and reach higher sizes, especially after week 10. This suggests a possible species effect, not just a climate effect. There is clear variation within each scenario and species. Measurements were taken regularly for all pears, so the design is balanced. Overall, Conference pears show larger sizes across all scenarios, and Scenario 1 leads to the highest growth in the later weeks.

3.1.6 Interaction Plot: Pear Size at Week 24

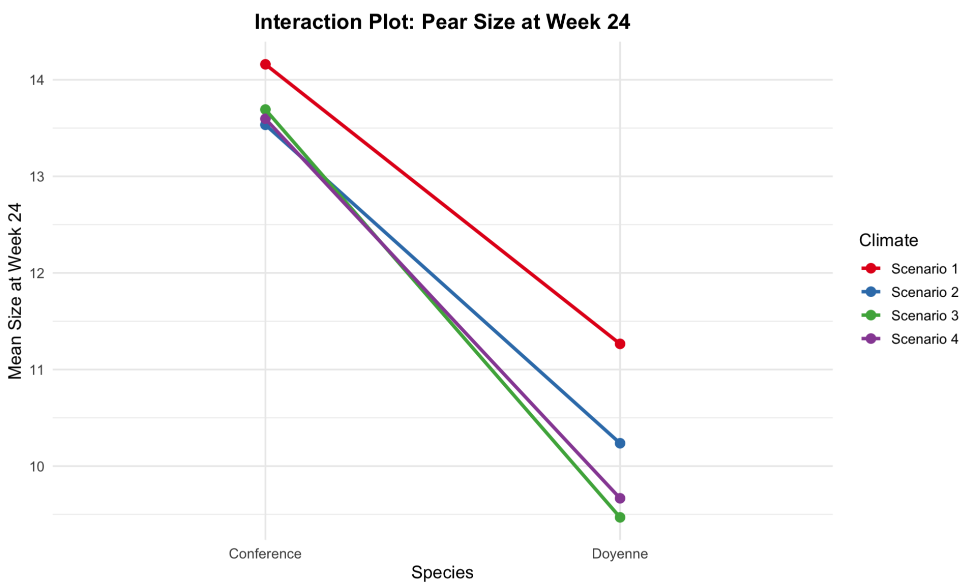

Figure 7: Interaction Plot: Pear Size at Week 24

Figure 7 compares mean pear sizes at week 24 for each species across scenarios.

A strong species effect is evident. Conference pears outgrow Doyenne consistently. No crossover indicates additive rather than interactive effects.

3.1.7 Distribution of Pear Size by Species and Climate at Selected Weeks

Figure 8 shows pear size distributions at weeks 0, 12, and 24 across species and climates.

At week 24, Scenario 1 leads to the highest size values, and Conference pears generally grow larger than Doyenne, regardless of climate. At week 0, All groups start with pear size = 0, as expected. At week 12, Conference pears are visibly larger than Doyenne pears across all scenarios. In scenario 1 (green) has the widest and highest violin for both species. Overall Conference pears consistently outgrow Doyenne across time and climates.

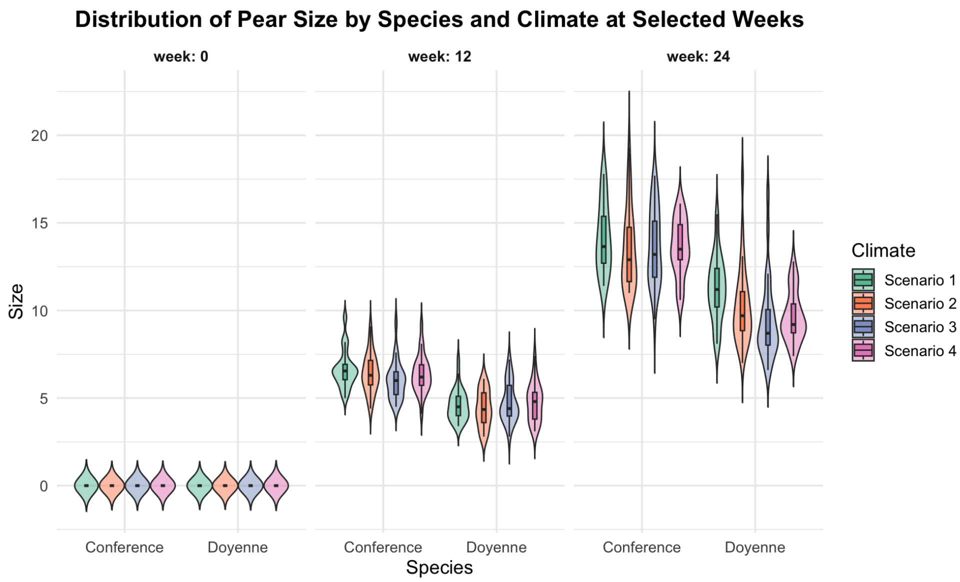

Figure 8: Distribution of Pear Size by Species and Climate at Selected Weeks

3.2 Primary Research Question: Effect of Climate and Species on Pear Quality Index

Based on the exploratory data analysis, it was evident that both climate and species could influence pear quality. However, the interaction between climate and species was tested and found to be statistically insignificant and thus was excluded from the final model.

A linear mixed-effects model was fitted using the continuous quality index as the outcome variable. Tree ID was included as a random effect, but the estimated variance was 0, indicating no detectable clustering at that level. The residual variance was estimated at 546.83.

Table 1: Solution for Fixed Effects (including SE)

Effect	**Level**	**Estimate**	**SE**	**95% CI**	**p-value**
Intercept	–	54.0123	1.6342	[50.5919, 57.4327]	<0.0001
Species					
Species	Doyenne	1.1083	1.5358	[-1.9057, 4.1223]	0.4707
Species	Conference (ref)	0	–	–	–
Climate					
Climate	Scenario 2	8.0835	2.1107	[3.9411, 12.2259]	0.0001
Climate	Scenario 3	5.2920	2.0840	[1.2022, 9.3819]	0.0113
Climate	Scenario 4	1.6595	2.2263	[-2.7097, 6.0287]	0.4562
Climate	Scenario 1 (ref)	0	–	–	–

Interpretation of Fixed Effects

The expected average quality index for conference pear under scenario1 is 54.01 (SE = 1.63), with a 95% confidence interval [50.59, 57.43], and is highly significant ($p < 0.0001$). Pears in Climate Scenario 2 are expected to have a quality index that is 8.08 points higher (SE = 2.11) than in Scenario 1, with a 95% confidence interval [3.94, 12.23], and the effect is highly significant ($p = 0.0001$).

The species effect for *Doyenne* (compared to Conference) is estimated at 1.11 (SE = 1.54), with a 95% confidence interval of [–1.91, 4.12], ($p = 0.4707$), Suggesting no strong evidence that species influences the quality index.

For the climate scenarios: **Scenario 2 (CO_2 removal) Scenario 3 (Sustainable energy)** shows Significant improvement over Scenario 1 and **Scenario 4 (Transportation)** shows no significance.

Conclusion: Scenario 2 yields the strongest positive effect on pear quality, followed by Scenario 3. Scenario 4 does not offer a meaningful improvement. The species variable is not a significant predictor of quality. The interaction term between species and climate was removed due to lack of significance in preliminary models.

Type III Tests of Fixed Effects

Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Species	1	915	0.52	0.4707
Climate	3	915	5.76	0.0007

Table 2: Type III Tests of Fixed Effects

Table 2 shows the Type III tests for the fixed effects. Climate had a statistically significant effect on the quality index ($F = 5.76$, $p = 0.0007$), while species did not ($F = 0.52$, $p = 0.4707$).

Pairwise Comparisons (Bonferroni Adjusted): Pairwise contrasts between scenarios Table 5 reveal that Scenario 2 significantly differs from Scenarios 1 and 4. Scenario 3 also shows significant improvement over Scenario 1.

Table 3: Pairwise Comparisons Between Climate Scenarios (Bonferroni Adjusted)

Comparison	Estimate	Std Error	p-value	Adj 95% CI	
				Lower	Upper
Scenario 2 vs Scenario 1	8.0835	2.1107	0.0008*	2.5028	13.6642
Scenario 2 vs Scenario 4	6.4240	2.2631	0.0278*	0.4405	12.4075
Scenario 3 vs Scenario 1	5.2920	2.0840	0.0676	-0.2179	10.8020
Scenario 2 vs Scenario 3	2.7914	2.1161	1.0000	-2.8035	8.3863
Scenario 3 vs Scenario 4	3.6326	2.2401	0.6314	-2.2904	9.5555
Scenario 4 vs Scenario 1	1.6595	2.2263	1.0000	-4.2268	7.5458

Significant at $\alpha = 0.05$ after Bonferroni correction.

Interpretation: Climate has a statistically significant impact on the pear quality index. Scenario 2 (CO_2 removal) leads to the highest quality, followed by Scenario 3 (sustainable energy). Scenario 1 (baseline) consistently results in the lowest mean quality. There is no significant difference in quality between pear species, and no interaction between climate and species was retained in the model due to non-significance.

3.3 Secondary Research Question: Effect of Climate and Species on Probability of Good Quality (Binary Outcome)

Exploratory analysis revealed strong visual patterns showing that climate interventions and pear species influence the likelihood of producing good-quality pears. Scenarios 2 (CO Removal) and 3 (Sustainable Energy) consistently led to higher and more stable quality yields, particularly for Conference pears. These patterns laid the foundation for the formal statistical modeling.

Table 4: GEE Model Results: Effect of Climate and Species on Good Quality Pears

Effect	Level	Estimate	SE	Lower 95% CI	Upper 95% CI	p-value
Intercept	–	-0.3358	0.1265	-0.5836	-0.0879	0.0079
Species						
Species	Conference	0.1393	0.1159	-0.0879	0.3665	0.2294
Species	Doyenne (ref)	0.0000	0.0000	0.0000	0.0000	–
Climate						
Climate	Scenario 2	0.4739	0.1586	0.1631	0.7847	0.0028
Climate	Scenario 3	0.4052	0.1557	0.1001	0.7103	0.0092
Climate	Scenario 4	0.0611	0.1686	-0.2694	0.3916	0.7170
Climate	Scenario 1 (ref)	0.0000	0.0000	0.0000	0.0000	–

Interpretation

- **Interaction effect:** Tested but found *not significant*, so it was excluded from the final model.
- **Species Effect (Conference vs. Doyenne):** Estimate = 0.139, $p = 0.229$
  This result is not statistically significant, suggesting no strong evidence that species influences pear quality in this model. Odds Ratio $\exp(0.139) \approx 1.15 \rightarrow$ Conference pears have about 15% higher odds of being good quality than Doyenne pears, though this is *not* significant. Compared to Scenario 1, Scenario 2 (CO_2 removal) showed significantly higher odds of producing good-quality pears, with an estimated 61% increase (OR $\approx$ 1.61, $p = 0.0028$). Scenario 3 (Sustainable energy) also had a significant effect, yielding about 50% higher odds (OR $\approx$ 1.50, $p = 0.0092$). In contrast, Scenario 4 (Transportation) showed no meaningful improvement (OR $\approx$ 1.06, $p = 0.7170$) and was not statistically significant.
- **Working Correlation:** Estimated within-tree correlation = $-0.0054 \rightarrow$ Shows very weak dependence among pears from the same tree. This justifies the use of a marginal model and indicates that within-tree clustering is negligible.

Conclusion from Statistical Modeling: The GEE results align with LMM Analysis. Climate interventions, especially **Scenario 2 (CO_2 removal)** and **Scenario 3 (sustainable energy)** have a significant and positive impact on the probability of pears being classified as good quality. While **Scenario 4** shows no statistically meaningful improvement. **Species**, although visually showing some trends, it is not statistically significant in this model.

3.4 Secondary Research Question: Effect of Climate and Species on Pear Growth Rate

Exploratory visualizations revealed distinct growth patterns in pear size over time, with Conference pears generally growing faster and climate scenario 1 unexpectedly showed faster late stage growth. These insights guided the inclusion of interaction terms between week, climate, and species.

The model focuses on pear size as a longitudinal continuous outcome measured over 24 weeks. Since the intercept starts at 0 for all pears (week 0), random intercepts were unnecessary. However, as we are interested in modeling changes in pear size over time (growth rate), we included random slopes for pear ID.

Table 5: Fixed Effects Estimates for Pear Size Growth Rate

Effect	**Level**	**Estimate**	**SE**	**95% CI**	**p-value**
week*climate	Scenario 2	0.3968	0.0071	[0.3829, 0.4106]	<0.0001
week*climate	Scenario 3	0.3870	0.0075	[0.3724, 0.4017]	<0.0001
week*climate	Scenario 4	0.4077	0.0076	[0.3928, 0.4226]	<0.0001
week*climate	Scenario 1 (ref)	0.4347	0.0074	[0.4203, 0.4491]	<0.0001
week*species	Conference	0.1466	0.0067	[0.1335, 0.1596]	<0.0001
week*species	Doyenne (ref)	0	–	–	–

Fixed Effects Estimates

Interpretation of Fixed Effects

The interaction between `week*climate*species` was ~~highly~~ non-significant and excluded from the final model for parsimony. All climate scenarios led to significant pear size growth ($p < 0.0001$). Scenario 1 resulted in the steepest growth rate (0.4347 cm/week), followed by Scenario 4 (0.4077cm/week), Scenario 2 (0.3968cm/week), and Scenario 3 (0.3870cm/week). Conference pears grew significantly faster than Doyenne, with an average additional growth of 0.1466 cm per week on average ($p < 0.0001$). **Residual Variance:** is 1.1977. **Random Slope Variance (pear ID)** is 0.002041 shows very low variability across individual growth rates.

Type III Tests of Fixed Effects

Table 6: Type III Tests of Fixed Effects

Effect	**Num DF**	**Den DF**	**F Value**	**p-value**
week*climate	3	4968	9.24	<0.0001
week*species	1	4968	481.78	<0.0001

Pairwise Comparisons Between Scenarios

Scenario 1 shows the steepest growth rate and significantly differs from all others. Scenario 3 grows significantly slower than both Scenario 1 and Scenario 4. The difference between Scenario 2 and 3 is not significant, despite Scenario 2 showing slightly higher growth.
Conclusion: The longitudinal modeling confirms the visual patterns from EDA. Pear growth rate is significantly influenced by both climate scenario and species. Conference pears grow faster than Doyenne. While Scenario 1 (no intervention) resulted in the highest growth. Climate interventions improve quality but not necessarily growth.

Table 7: Pairwise Contrasts of Week*Climate Effects

Comparison	Estimate	SE	t-value	p-value
Scenario 2 – Scenario 1	–0.0379	0.0093	–4.06	<0.0001
Scenario 3 – Scenario 1	–0.0476	0.0096	–4.97	<0.0001
Scenario 4 – Scenario 1	–0.0270	0.0096	–2.80	0.0052
Scenario 2 – Scenario 3	0.0097	0.0092	1.05	0.2915
Scenario 2 – Scenario 4	–0.0110	0.0093	–1.18	0.2374
Scenario 3 – Scenario 4	–0.0207	0.0096	–2.17	0.0299

3.5 Exploring soil condition features with a multivariate method

3.5.1 Principal Component Analysis (PCA) Results

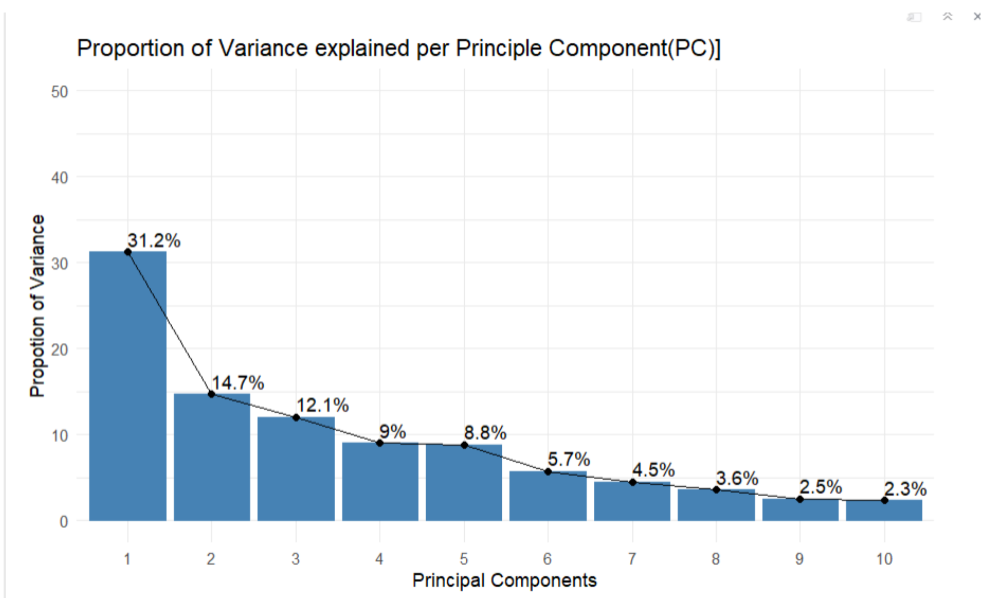

Figure 9: Variance Explained in Principal Components

Figure 9 illustrates the proportion of variance explained by each principal component. The x-axis represents the principal components ordered by their explained variance. The y-axis shows the percentage of explained variance. Some principal components exhibit near-zero variance due to multicollinearity. A line with dots visualizes the trend in explained variance across components. Although there is no clear "elbow" point in the plot, we retained the first two principal components (PC1 and PC2) for visualization, as they together explain approximately 46% of the total variance.

3.5.2 Biplot Analysis of Climate Scenarios

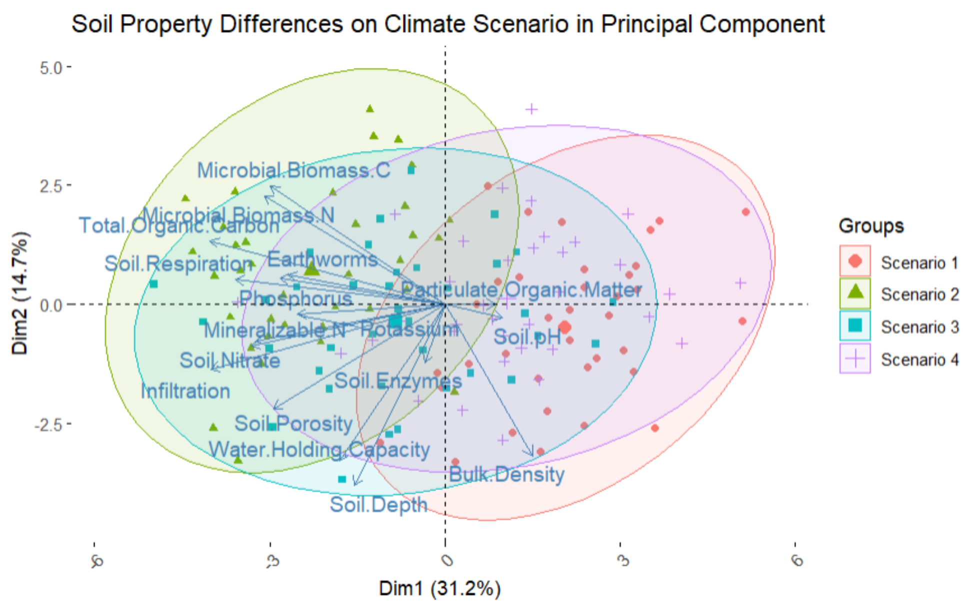

Figure 10: Soil features association with Climate Scenarios

Figure 10 presents the PCA biplot visualized the first two principal components of the 17 soil variables, where each point represents a soil sample colored by climate scenario. Arrows indicate variable contributions and correlations to the principal components, while ellipses denote 95% confidence regions.

From the plot, it appears that Scenario 1 and Scenario 4 mostly clustered on the right side (positive PC1 values), while Scenarios 2 and 3 were positioned more on the left (negative PC1 values). This distribution indicates that PC1 captures the distinction between the climate scenarios, grouping Scenario 1 and 4 together, and Scenario 2 and 3 into a separate cluster.

The comparison of soil characteristics across the four climate scenarios was assessed by projecting the data points (samples) orthogonally onto the feature vectors (arrows). For example, when projecting the samples from Scenarios 1 and 4 onto the `Bulk.Density` vector, the resulting values were relatively higher than average, indicating that soils in these two scenarios tended to have greater bulk density compared to Scenarios 2 and 3.

In contrast, projections of samples onto vectors such as `Organic.Carbon`, `Microbial.Biomass.C`, `Microbial.Biomass.N`, and `Soil.Respiration` showed higher values for Scenarios 2 and 3. This suggests that soils under Scenarios 2 and 3 were generally more biologically active and richer in organic matter than those in Scenarios 1 and 4.

3.6 Multivariate Confirmation of Climate Effects on Soil and Pear Production

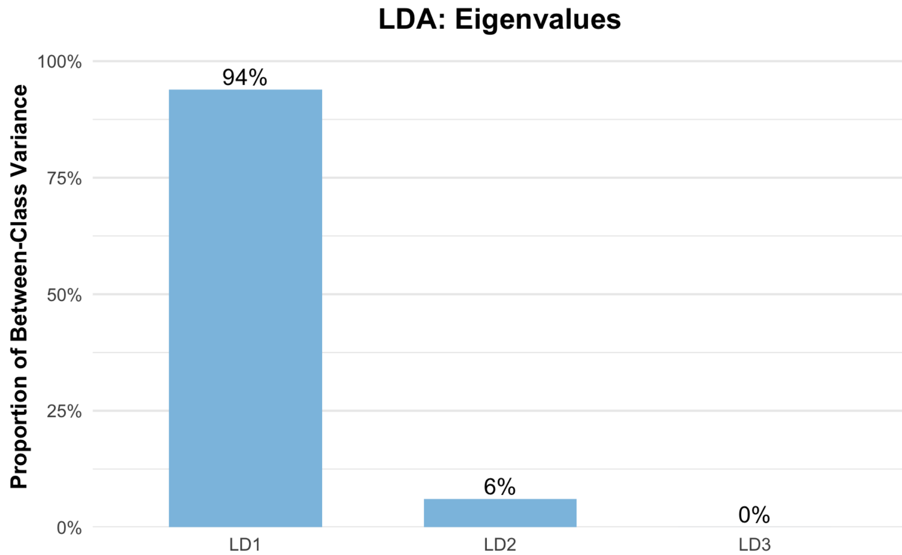

Figure 11: LDA eigenvalue barplot showing proportion of between-class variance explained by LD1, LD2, and LD3.

The eigenvalue barplot figure 11 revealed that the first linear discriminant (LD1) captured approximately 94% of the between-group variance, while LD2 accounted for 6%. This shows that the majority of climate-based separation can be captured in a one-dimensional space.

The LDA scatterplot in Figure 12 showed partial separation among the four climate scenarios along LD1 and LD2, with considerable overlap, especially between Scenarios 2, 3, and 4. However, Scenario 1 formed a slightly more distinct cluster more aligned towards LDA 2. A closer view of how soil features and pear numbers contributed to the group separation could be found in the next part.

The variable loadings on LD1 indicated that climate scenarios 2, 3 and 4 were associated with higher infiltration, soil depth, potassium, and phosphorus and lower water holding capacity, soil nitrate, and mineralizable nitrogen.

The contribution of the number of pears to both LD1 and LD2 was nearly zero, indicating that its contribution to the discrimination between climate scenarios was not much. This suggested that the observed separation was driven primarily by differences in soil conditions rather than direct differences in pear yield. Hence, there was not enough evidence of the impact of climate change on pear yield.

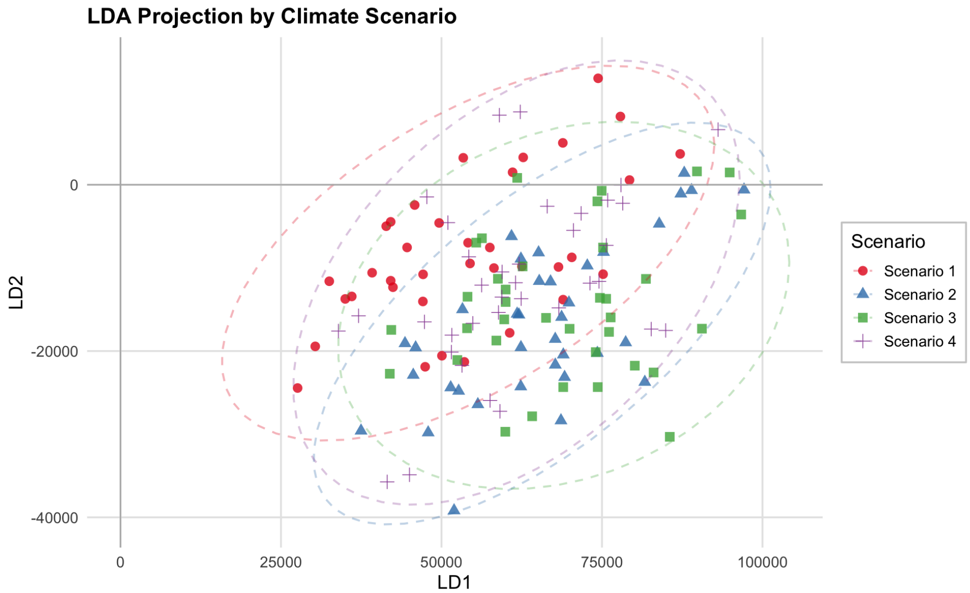

Figure 12: LDA scatterplot of LD1 vs LD2, showing the separation between different climate scenarios.

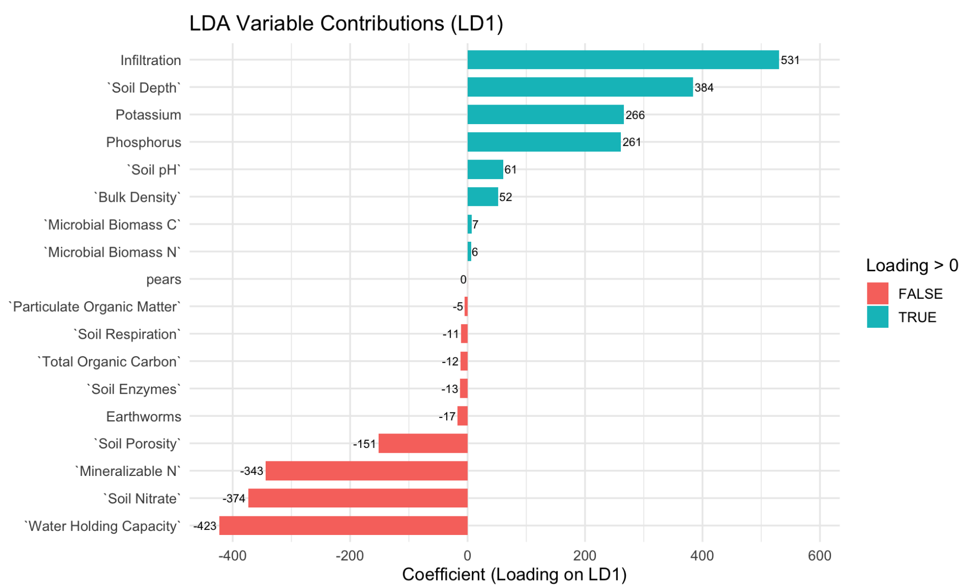

Figure 13: Variable contributions to LD1: positive and negative loadings indicating importance in discriminating climate scenarios.

4 Discussion

This project aimed to assess the impact of climate interventions and species on pear quality, growth, and soil health. Our analysis revealed that Climate Scenario 2 had the most pronounced effect on both the pear quality score and the associated quality indicators among the four groups. In contrast, the longitudinal analysis of weekly size measurements indicated that Scenario 1 was linked to the fastest average rate of size development. This suggests that rapid growth in pear size does not necessarily correspond to improved quality. Rather, slower and more stable growth patterns, as observed in Scenario 2, were associated with higher quality outcomes in this study. The unsupervised PCA (without taking the prior grouping information into account) showed that soil samples between scenario 1 & 4 and scenarios 2 & 3 could be separated. Further, LDA demonstrated moderate separation between Scenario 1 and the remaining scenarios. Interestingly, the finding showed no difference in pear yield among 4 scenarios. While the analysis did not reveal substantial differences in pear yield across scenarios, it is important to note that LDA was used as an exploratory tool, and additional confirmatory analysis would be necessary to draw firm conclusions.

References

[1] Marques, I., Ramalho, J.C., and Ribeiro-Barros, A.I. (2023). Plant responses to climate change. *International Journal of Molecular Sciences*, **24(21)**, 15902.

[2] ESS Team. (2025). How climate change is reshaping pear production worldwide. *Unpublished report.*

Appendix - R/SAS code

This section contains only the core code. it omits the exploratory data analysis and plotting scripts and includes solely the statistical modeling routines, some part of the PCA and LDA components.

SAS Code:

```
* Model for primary research question;
proc mixed data=datasets.finaldata;
class climate (ref='Scenario 1') species (ref='Conference') tree_id;
model quality_index = species climate / solution cl;
random tree_id;
lsmeans climate / diff adjust=bon cl;
run;
```

```
* Model for Question a;
proc genmod data=pears_binary descending;
class species climate(ref='Scenario 1') tree_id;
model quality_b = species climate / dist=bin link=logit;
repeated subject=tree_id / type=exch corrw modelse;
run;

* Model for Question b;
proc mixed data=datasets.peargrowthdata;
class pear_id climate (ref = "Scenario 1") species (ref = "Doyenne");
model size = climate*week species*week /noint solution cl;
random week/ type=ar(1) subject= pear_id;
estimate "Sce2 - Sce1" climate*week 1 0 0 -1;
estimate "Sce3 - Sce1" climate*week 0 1 0 -1;
estimate "Sce4 - Sce1" climate*week 0 0 1 -1;
estimate "Sce2 - Sce3" climate*week 1 -1 0 0;
estimate "Sce2 - Sce4" climate*week 1 0 -1 0;
estimate "Sce3 - Sce4" climate*week 0 1 -1 0;
run;
```

R Code:

```
# PCA Code

#standardizing the extracted data because there are of different units
soil_data_numeric_only_scaled<- scale(soil_data_numeric_only) # Centers and scales the data
head(soil_data_numeric_only_scaled)
#running PCA
pca_soil_data<- prcomp(soil_data_numeric_only_scaled, center = FALSE,scale. = FALSE)
summary(pca_soil_data)
# Create the scree plot
fviz_eig(pca_soil_data,
         addlabels = TRUE,
         ylim = c(0, 50))  # Adjust the y-axis limit if needed

# PCA biplot colored by climate
fviz_pca_biplot(pca_soil_data,
                label = "var", # Show variable names
                habillage = soil_data$climate, # Color by climate
                addEllipses = TRUE, # Add confidence ellipses
```

```
                ellipse.level = 0.95, # 95% confidence ellipses
                repel = TRUE) # Avoid text overlapping
# PCA biplot colored by species
fviz_pca_biplot(pca_soil_data,
                label = "var", # Show variable names (soil variables)
                habillage = soil_data$species, # Color by species
                addEllipses = TRUE, # Add group ellipses
                ellipse.level = 0.95, # 95% confidence level
                repel = TRUE) # Avoid overlapping text

# LDA Code
# Load required libraries
library(MASS)
library(dplyr)

# Step 1: Use your data (assuming it's in a data frame called `lda_data`)
# Convert climate to factor
soil_data$climate <- as.factor(soil_data$climate)

# Step 2: Select numeric predictors (soil + pears)
predictors <- soil_data %>%
  dplyr::select(where(is.numeric))  # This includes all 17 soil vars + pears
# Step 3: Remove rows with missing values (or impute if needed)
# predictors_clean <- predictors %>%
#   mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))
predictors<- predictors %>%
  mutate(across(everything(), ~ scale(.)))

# Step 4: Fit LDA model
lda_model <- lda(soil_data$climate ~ ., data = predictors)
```