

Determining Optimal Simulation Repetition for Forest Stand Modeling: A Replication Analysis of the FORCEPS Forest Growth Model

Clementine de Montgolfier

2025-08-26

Abstract

Forest growth modeling is essential for sustainable forest management and climate change mitigation strategies. However, determining the optimal number of simulation repetitions required to achieve statistically reliable estimates is a challenge in computational forest ecology. This study addresses this question for the FORCEPS (Forest Ecosystem Simulation Platform) forest growth model using data from the Retz forest in France.

We conducted a comprehensive replication analysis using 5,000 individual FORCEPS simulations to determine the minimum number of simulations (n) required for stable estimates of key forest metrics including basal area, volume, biomass, productivity, and mortality. Our methodology employed a two-stage approach: first, we assessed the convergence of the coefficient of variation (CV) across different numbers of repetitions (m); second, we analyzed how CV and standardized error decrease with increasing simulation numbers (n).

Results demonstrate that 10 repetitions ($m = 10$) are sufficient for coefficient of variation convergence across all tested forest metrics. For simulation numbers, our analysis reveals that 10 simulations ($n = 10$) provide adequate precision for most forest management applications. While 50 simulations offer superior precision, the marginal improvement does not justify the increased computational cost for our applications.

Introduction

Study Objectives

This study aims to establish optimal simulation parameters for the FORCEPS forest model through a comprehensive replication analysis. Specifically, we address two fundamental questions:

1. **Repetition optimization (m):** What is the minimum number of repetitions required to achieve stable coefficient of variation estimates?
2. **Simulation optimization (n):** How many individual simulations are needed to achieve acceptable precision in forest metric estimates?

Study Site and Model Description

Study Area

The analysis focuses on the Retz forest (patch RETZ_00137_02), located in northern France.

FORCEPS Model

FORCEPS is a spatially explicit, individual-based forest model that simulates tree growth, competition, mortality, and regeneration processes. The model operates at the individual tree level and incorporates

climate variables, soil characteristics, and species-specific physiological parameters. Some output variables include:

- Total basal area (m²/ha)
- Total volume (m³/ha)
- Total biomass (t/ha)
- Annual productivity (m²/ha/year)
- Tree mortality (individuals/ha/year)

Methods

Simulation Framework

Data Generation

We conducted 5,000 independent FORCEPS simulations for the Retz forest patch, each spanning 80 years (2020-2100) with annual time steps. All simulations used identical initial forest conditions but different random seeds to capture stochastic variability in growth processes. The simulation protocol included:

- **Patch area:** 1,000 hectares
- **Simulation period:** 80 years (2020-2100)
- **Climate scenario:** Current climate conditions (retz_act.climate)
- **Management scenario:** 80_3_1_25_FSy1-80 (representing typical silvicultural practices)
- **Species composition:** 9 potential species based on local forest inventory data

Statistical Analysis Framework

Our analysis employed a two-stage approach to optimize both repetition numbers (m) and simulation numbers (n):

Stage 1: Repetition Analysis (m optimization) For a fixed number of simulations (n), we assessed how the coefficient of variation (CV) stabilizes as the number of repetitions (m) increases. Each repetition represents an independent sampling of n simulations from our 5,000-simulation pool.

Stage 2: Simulation Analysis (n optimization)

Using the optimal repetition number determined in Stage 1, we analyzed how CV and standardized error decrease as the number of simulations (n) increases.

Statistical Metrics

Coefficient of Variation (CV)

The coefficient of variation was calculated as:

$$CV = \frac{\sigma}{\mu}$$

where sigma is the standard deviation and mu is the mean of the forest metric across repetitions.

Standardized Error

The standardized error was calculated as:

$$SE = \frac{1.833 \times \sigma}{\sqrt{n}}$$

where 1.833 represents the 90% confidence interval multiplier for a t-distribution, sigma is the standard deviation, and n is the number of simulations.

Convergence Criteria

We used the following convergence criteria: - **Error thresholds:** Standardized error < 5% and < 10% of mean values

Results

Repetition Optimization Analysis (Stage 1)

The first stage of our analysis focused on determining the minimum number of repetitions (m) required for stable coefficient of variation estimates. We tested repetition numbers ranging from 5 to 400 across three different simulation group sizes (n = 2, 10, 50 i.e. each repetition is a mean of n simulations) for three key forest metrics.

Analytical Approach

For each combination of simulation number (n) and repetition number (m), we:

1. Randomly sampled n simulations from our 5,000-simulation pool
2. Calculated mean values for each forest metric on the n simulations
3. Repeated this process m times to generate m repetitions
4. Computed the coefficient of variation across these m repetitions
5. Assessed CV stability as m increased

Results for Total Basal Area

Low Simulation Number (n = 2) Analysis: With only 2 simulations per repetition, the coefficient of variation remains relatively high (approximately 0.10) and requires approximately 50 repetitions for stabilization. The standard error is substantial (4 m²/ha), indicating significant uncertainty in estimates.

Moderate Simulation Number (n = 10) Analysis: With 10 simulations per repetition, the coefficient of variation decreases substantially to approximately 0.05 and achieves stable convergence at just 5-10 repetitions. The standard error reduces to 1 m²/ha, representing a significant improvement in precision.

High Simulation Number (n = 50) Analysis: With 50 simulations per repetition, the coefficient of variation achieves excellent stability (0.02) immediately, with minimal variation across repetition groups. The standard error is very low (0.2 m²/ha), indicating highly precise estimates.

Results for Total Volume

Low Simulation Number (n = 2)

Moderate Simulation Number (n = 10)

Results for Total Biomass

Low Simulation Number (n = 2)

Moderate Simulation Number (n = 10)

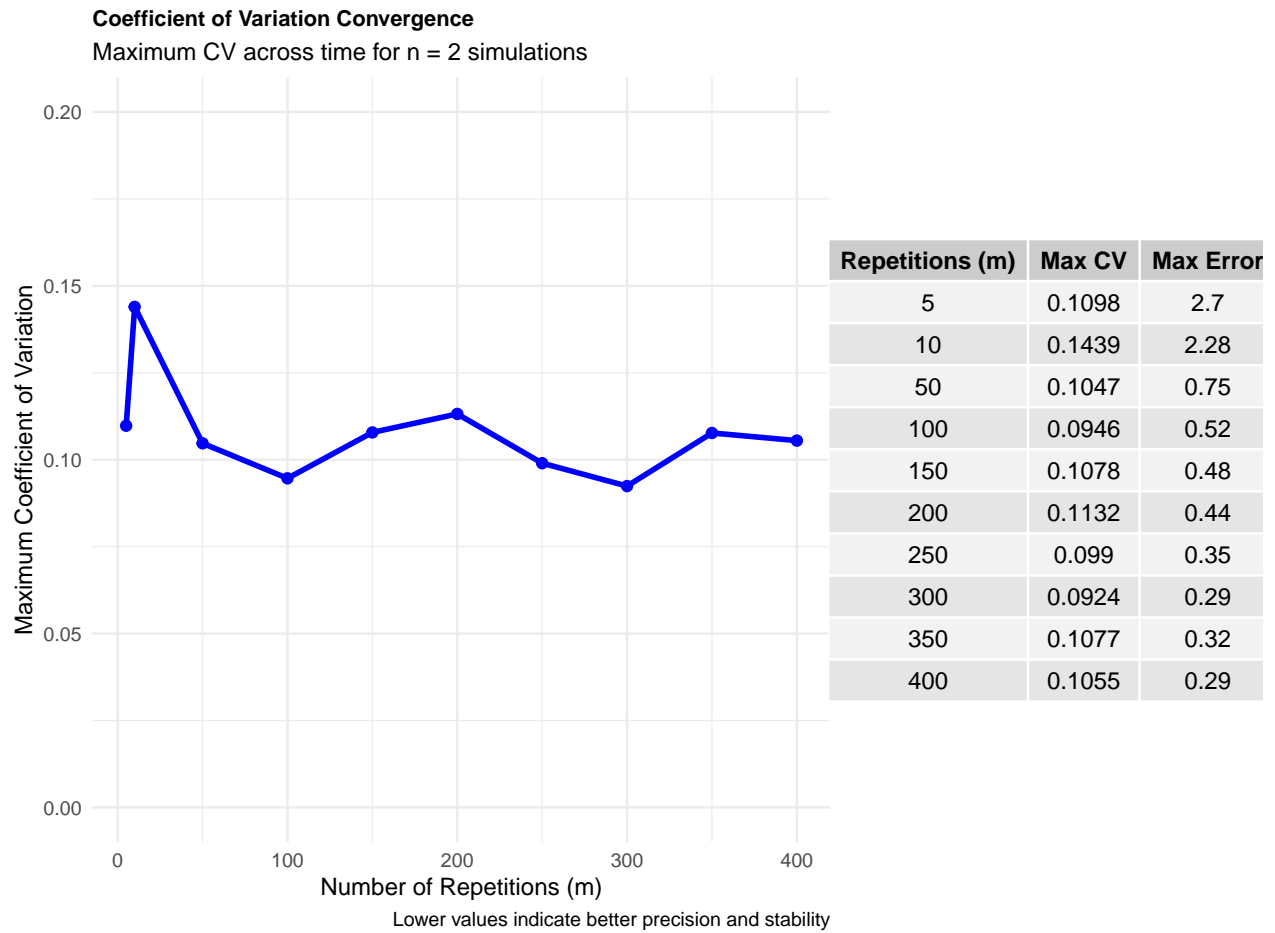


Figure 1: Figure 1: Coefficient of variation analysis for total basal area with 2 simulations per repetition. The left panel shows individual repetition trajectories across different repetition group sizes. The right panel demonstrates CV convergence, with values stabilizing around 0.10 after 50 repetitions.

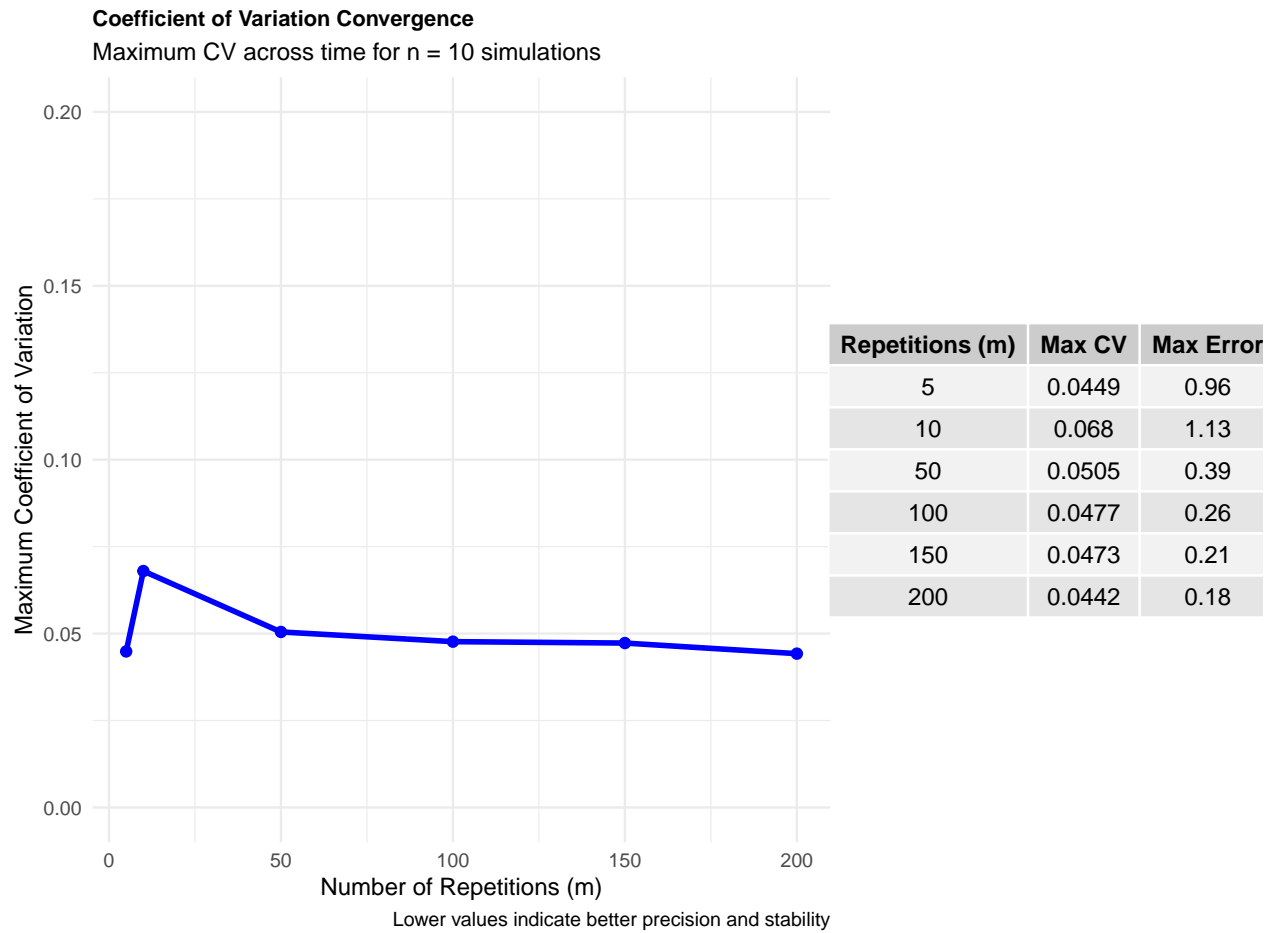


Figure 2: Figure 2: Coefficient of variation analysis for total basal area with 10 simulations per repetition. CV values stabilize around 0.05 with good convergence achieved at 10 repetitions, representing an optimal balance between precision and computational efficiency.

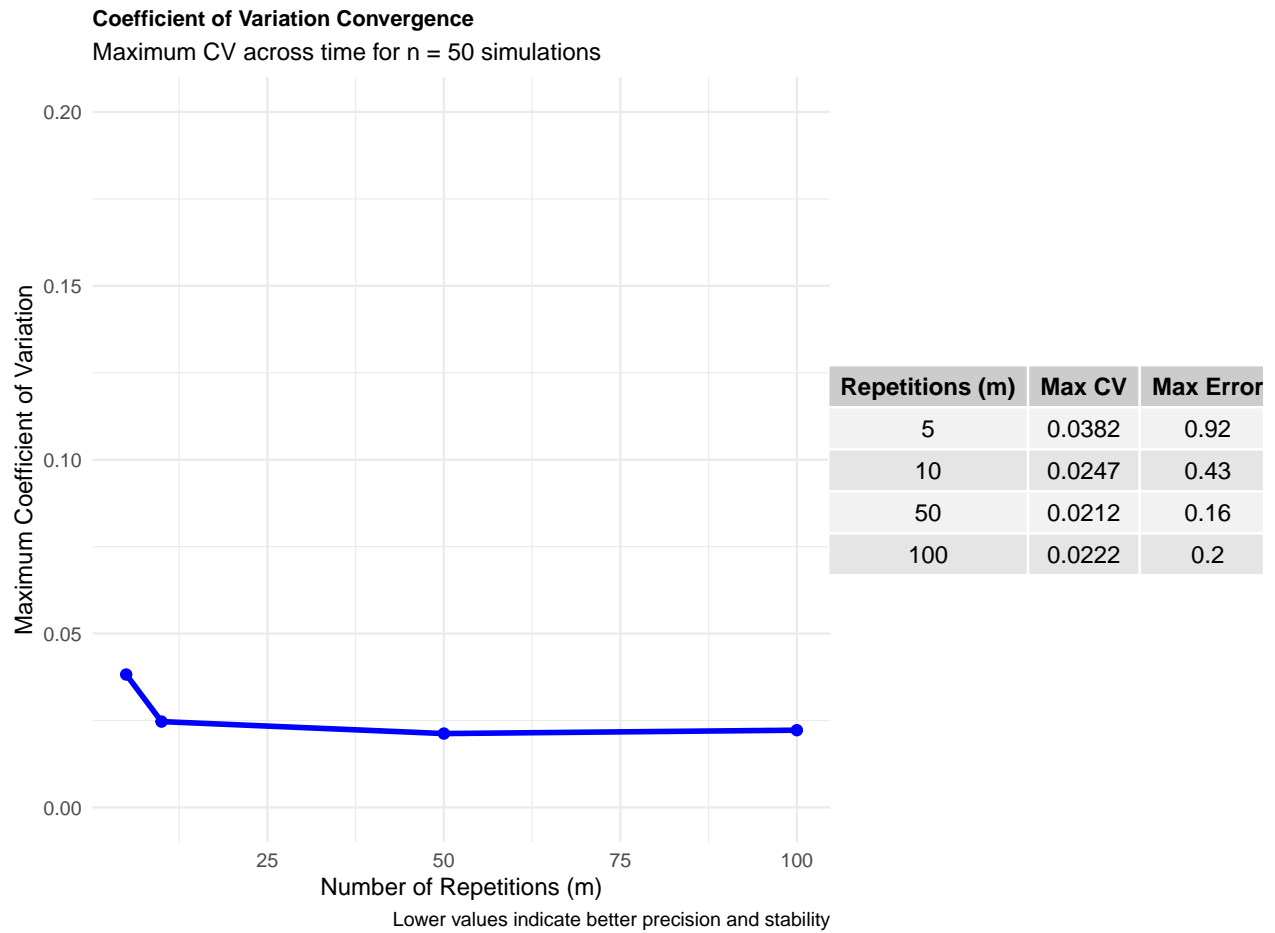


Figure 3: Figure 3: Coefficient of variation analysis for total basal area with 50 simulations per repetition. CV values are very low (approximately 0.02) and demonstrate excellent stability from the first repetition group, indicating high precision with minimal uncertainty.

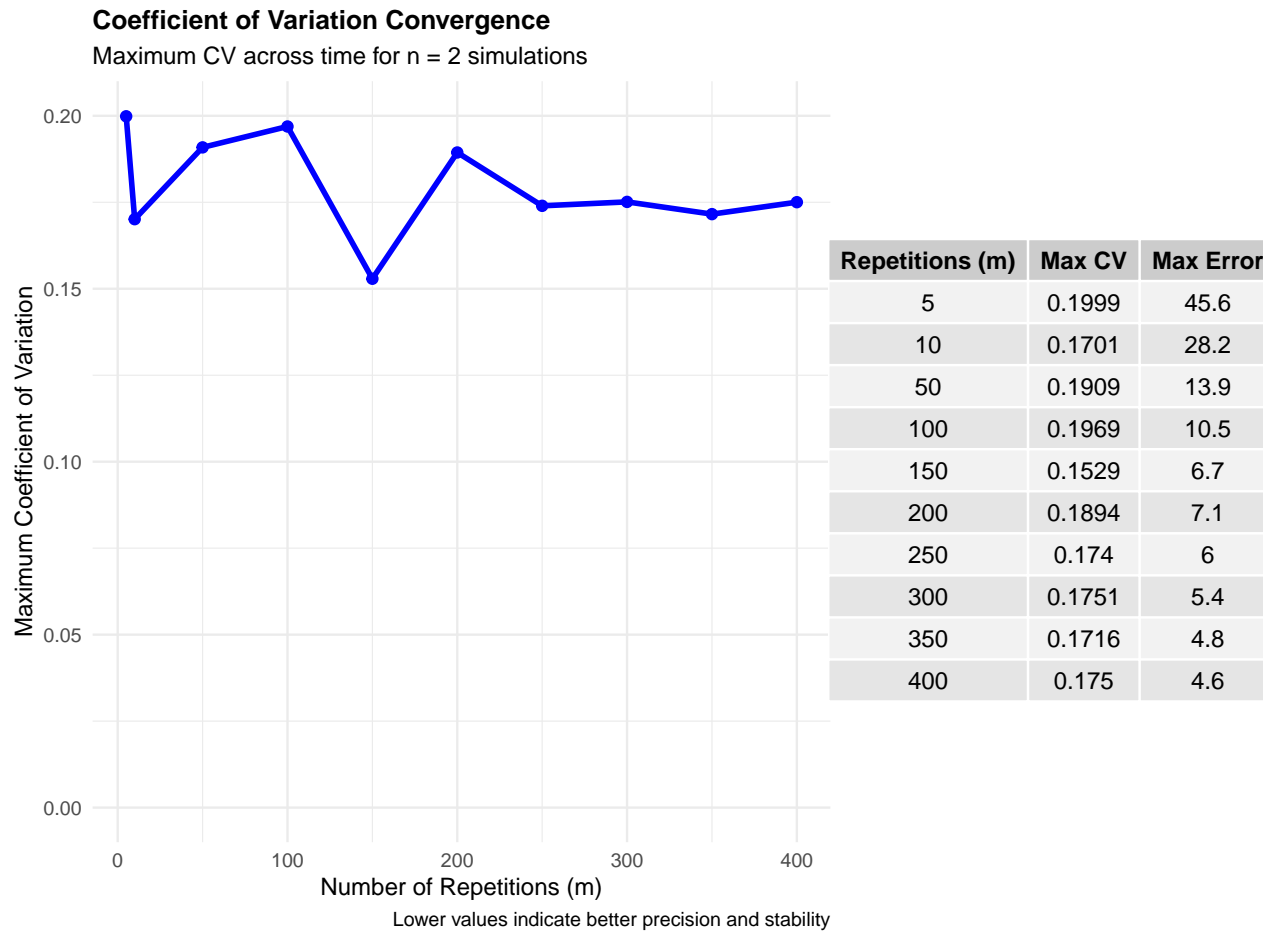


Figure 4: Figure 4: Coefficient of variation analysis for total volume with 2 simulations per repetition. Similar patterns to basal area are observed, with CV values requiring substantial repetitions for convergence.

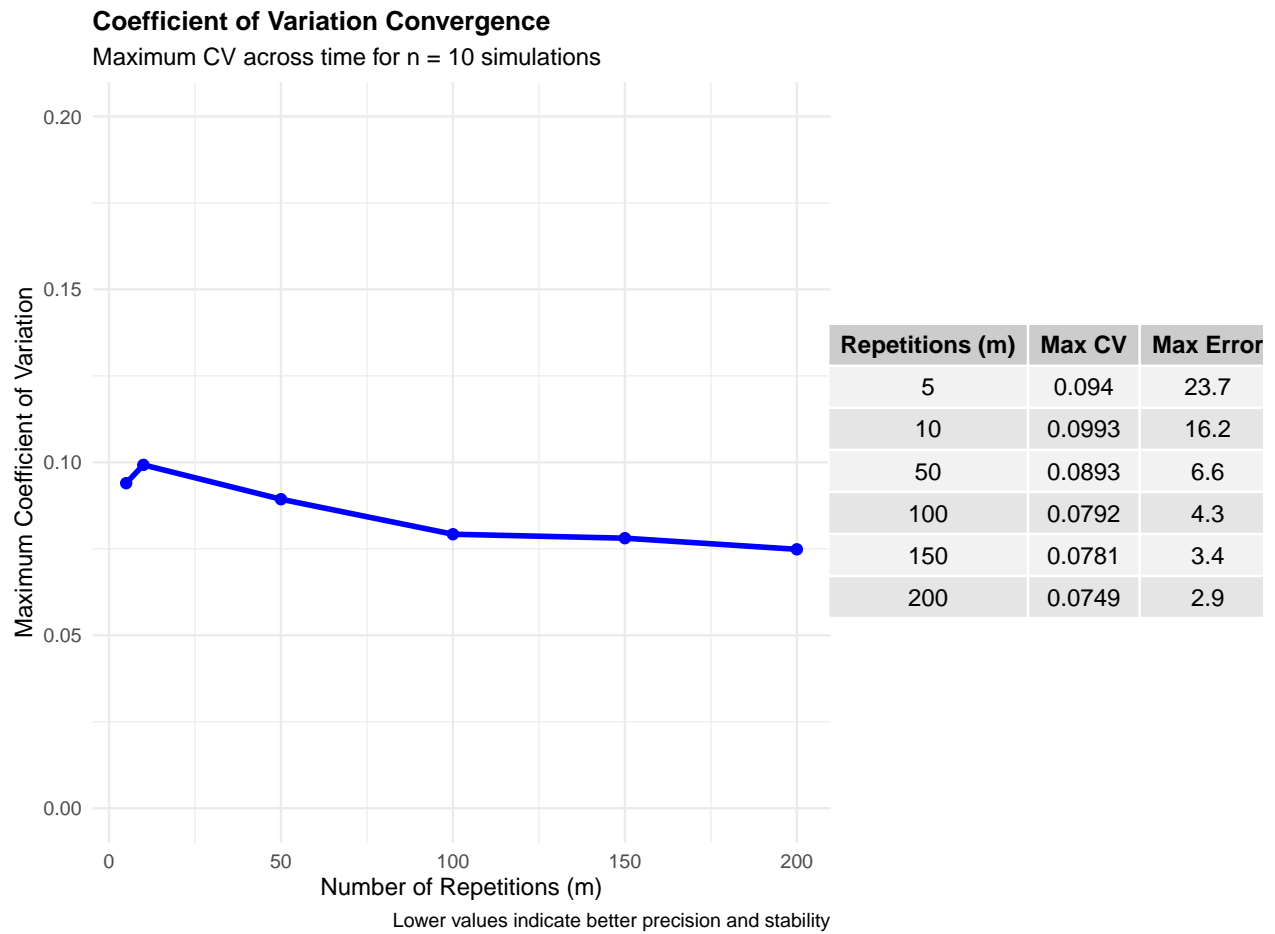


Figure 5: Figure 5: Coefficient of variation analysis for total volume with 10 simulations per repetition. Convergence patterns mirror those observed for basal area, confirming the robustness of our findings across forest metrics.

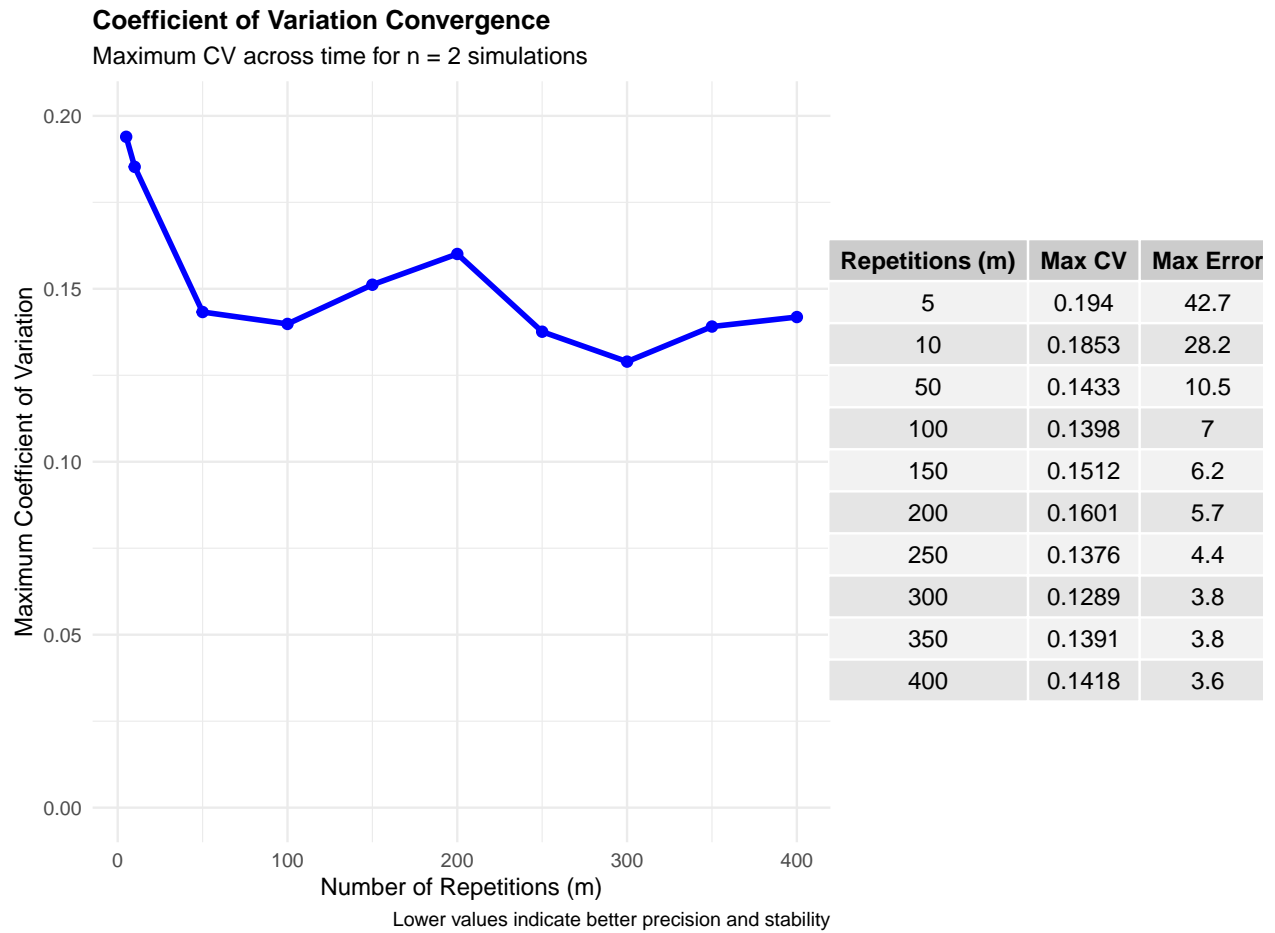


Figure 6: Figure 6: Coefficient of variation analysis for total biomass with 2 simulations per repetition. Biomass estimates show similar convergence patterns to other forest metrics, validating our methodological approach.

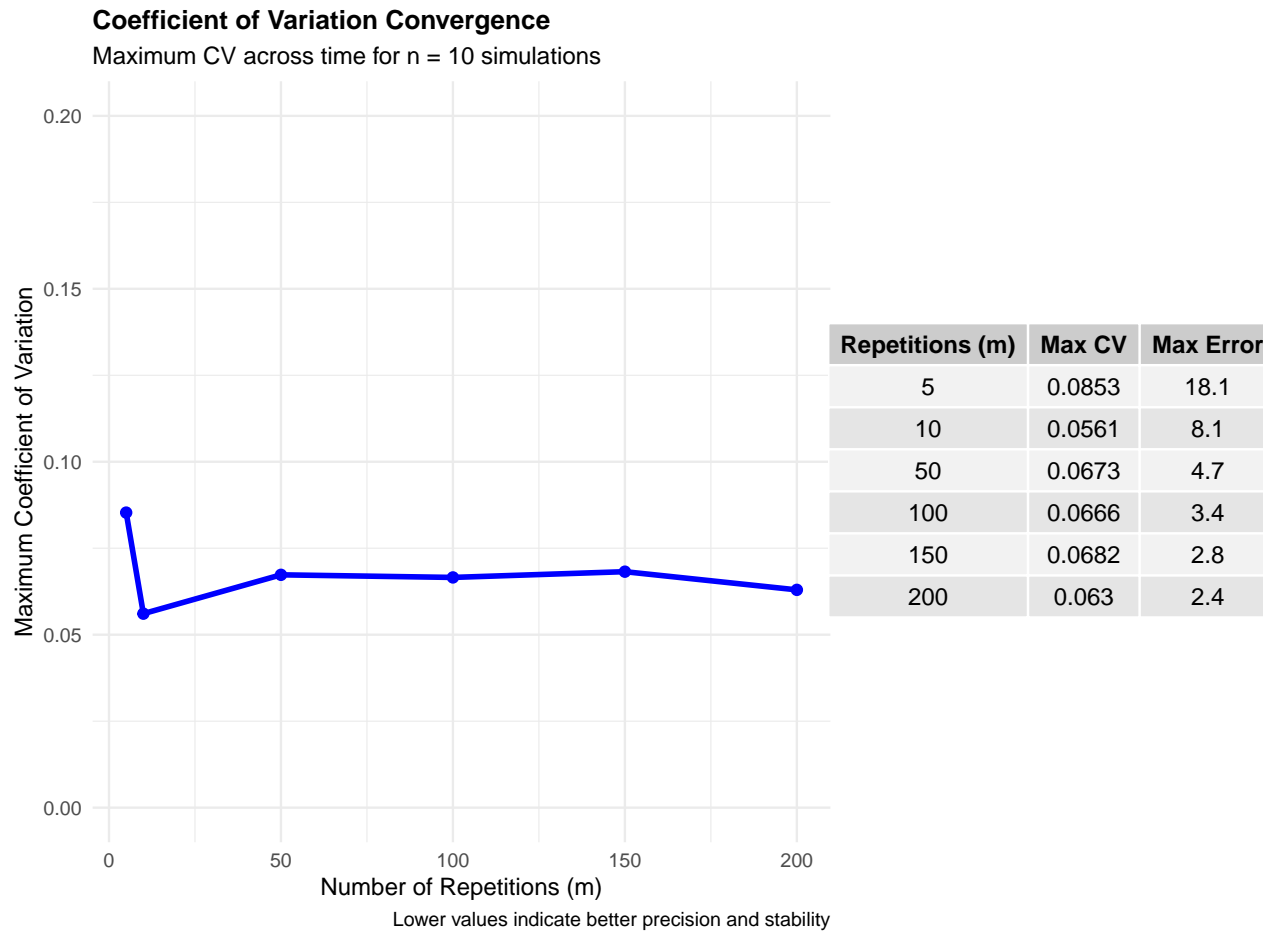


Figure 7: Figure 7: Coefficient of variation analysis for total biomass with 10 simulations per repetition. Results consistently support the conclusion that 10 repetitions provide adequate convergence for forest modeling applications.

Stage 1 Conclusions

The analysis demonstrates that while 50 repetitions provide superior precision from the outset, 10 repetitions achieve acceptable precision (if we use the mean of at least 10 simulations), establishing our framework for Stage 2 analysis.

Simulation Optimization Analysis (Stage 2)

Having established that 10 repetitions provide optimal convergence ($m = 10$), we now focus on determining the minimum number of simulations (n) required to achieve acceptable precision levels. This analysis examines how coefficient of variation and standardized error decrease as simulation numbers increase from 1 to 150.

Analytical Framework

Using $m = 10$ repetitions, we tested simulation group sizes ranging from 1 to 150 simulations. For each simulation group size, we:

1. Generated 10 independent repetitions, each containing n simulations
2. Calculated mean forest metrics for each repetition
3. Computed coefficient of variation and standardized error across repetitions
4. Assessed precision improvement with increasing simulation numbers

Results for Total Basal Area

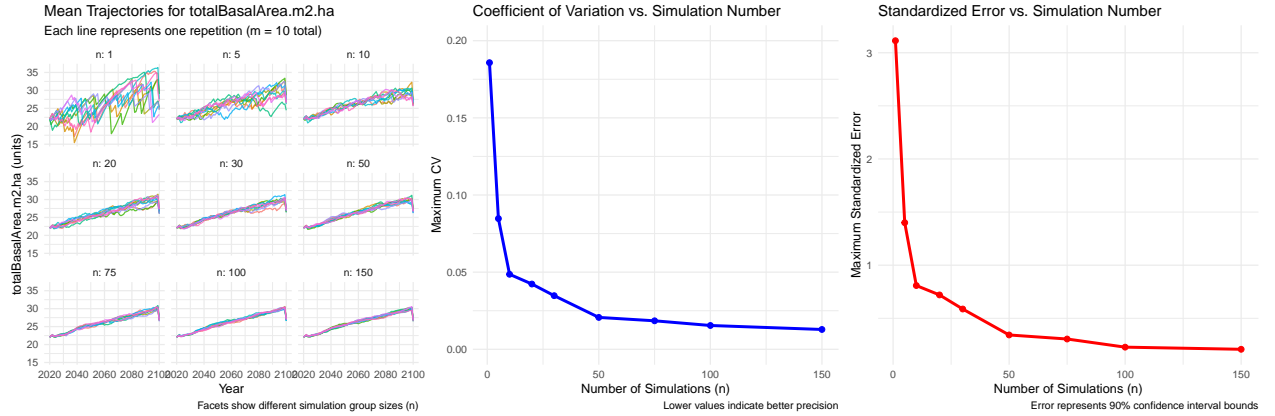


Figure 8: Figure 8: Simulation optimization analysis for total basal area. Left panel shows mean trajectories across different simulation group sizes. Center panel demonstrates CV reduction with increasing simulation numbers. Right panel shows standardized error convergence, with significant improvements observed up to $n=20$.

Table 1: Table 1: Precision metrics for total basal area across simulation numbers

Simulations (n)	Max CV	Max Error	Mean CV	Mean Error
1	0.1858	3.11	0.1003	1.54
5	0.0847	1.40	0.0493	0.77
10	0.0486	0.81	0.0272	0.42
20	0.0423	0.72	0.0246	0.38
30	0.0347	0.59	0.0169	0.26
50	0.0207	0.34	0.0144	0.22
75	0.0185	0.31	0.0118	0.18
100	0.0154	0.23	0.0086	0.13

Simulations (n)	Max CV	Max Error	Mean CV	Mean Error
150	0.0128	0.21	0.0084	0.13

Results for Total Volume

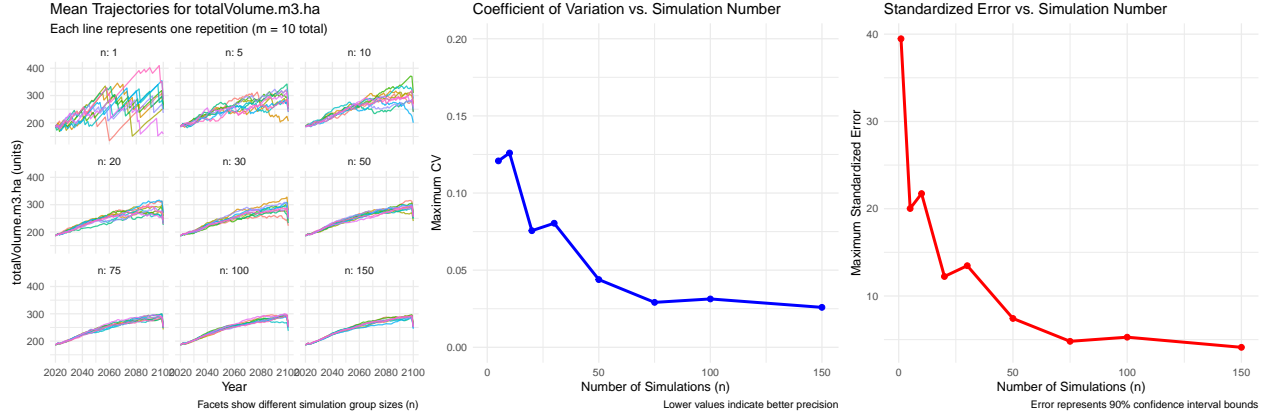


Figure 9: Figure 9: Simulation optimization analysis for total volume. Volume estimates show similar convergence patterns to basal area, with optimal precision achieved around $n=10-20$ simulations.

Table 2: Table 2: Precision metrics for total volume across simulation numbers

Simulations (n)	Max CV	Max Error	Mean CV	Mean Error
1	0.2430	39.5	0.1450	21.9
5	0.1208	20.0	0.0649	9.8
10	0.1260	21.7	0.0544	8.5
20	0.0756	12.2	0.0395	5.9
30	0.0804	13.5	0.0353	5.4
50	0.0439	7.4	0.0233	3.5
75	0.0291	4.8	0.0180	2.7
100	0.0313	5.3	0.0179	2.7
150	0.0259	4.1	0.0145	2.2

Results for Total Biomass

Table 3: Table 3: Precision metrics for total biomass across simulation numbers

Simulations (n)	Max CV	Max Error	Mean CV	Mean Error
1	0.2187	33.3	0.1241	18.4
5	0.1283	20.5	0.0614	9.2
10	0.0606	9.3	0.0359	5.2
20	0.0385	5.7	0.0242	3.4
30	0.0457	7.2	0.0269	3.9
50	0.0377	5.9	0.0186	2.7
75	0.0198	3.1	0.0129	1.9
100	0.0235	3.6	0.0133	1.9
150	0.0143	2.3	0.0094	1.3

Simulations (n)	Max CV	Max Error	Mean CV	Mean Error
-----------------	--------	-----------	---------	------------

Analysis of Productivity Metrics

Working on productivity analysis to: - Obtain more temporally stable values (less variation, making error less date-dependent) - Reduce dependence on initial values (strong divergence in early years affects other accumulated metrics)

Species-Level Productivity Analysis

Species-Level Analysis Notes: Coefficient of variation is unreliable for some species because certain values are zero (making the mean null and CV infinite). Additionally, some species have very low productivity, resulting in very high coefficient of variation values.

Total Productivity Analysis

Threshold Analysis Results: - For 5% error threshold: **50 simulations** required
- For 10% error threshold: **10 simulations** required

Mortality Analysis

Analyzing tree mortality (deadNumber_ha) from the productivity scene dataset.

Mortality Analysis Results: - n = **10** for 10.5% error - n = **20** for 10% error
- n = **50** for 5% error

Precision Gained by Increasing Repetition Numbers

To further validate our findings, we conducted an additional analysis examining precision improvements with varying repetition numbers while maintaining fixed simulation group sizes.

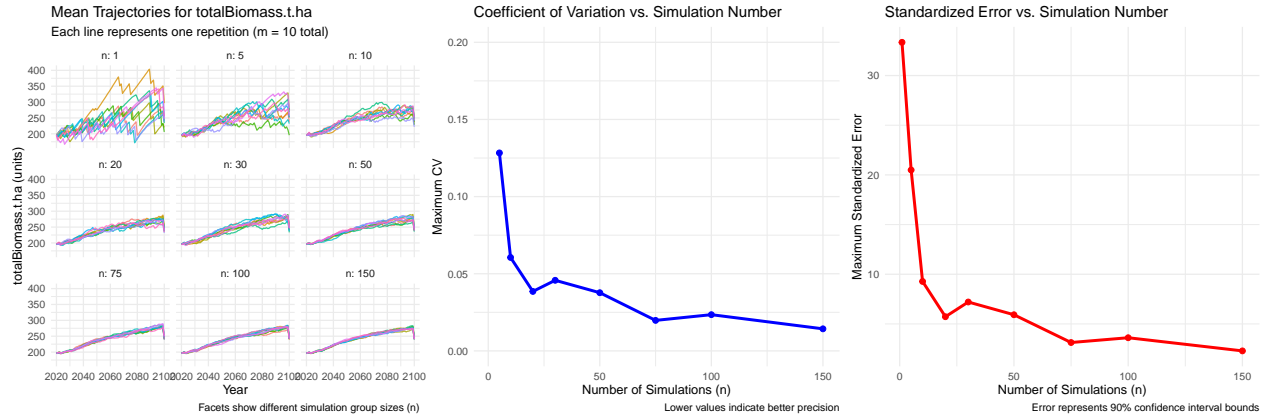


Figure 10: Figure 10: Simulation optimization analysis for total biomass. Biomass estimates demonstrate consistent convergence behavior, supporting the robustness of our optimization approach across multiple forest metrics.

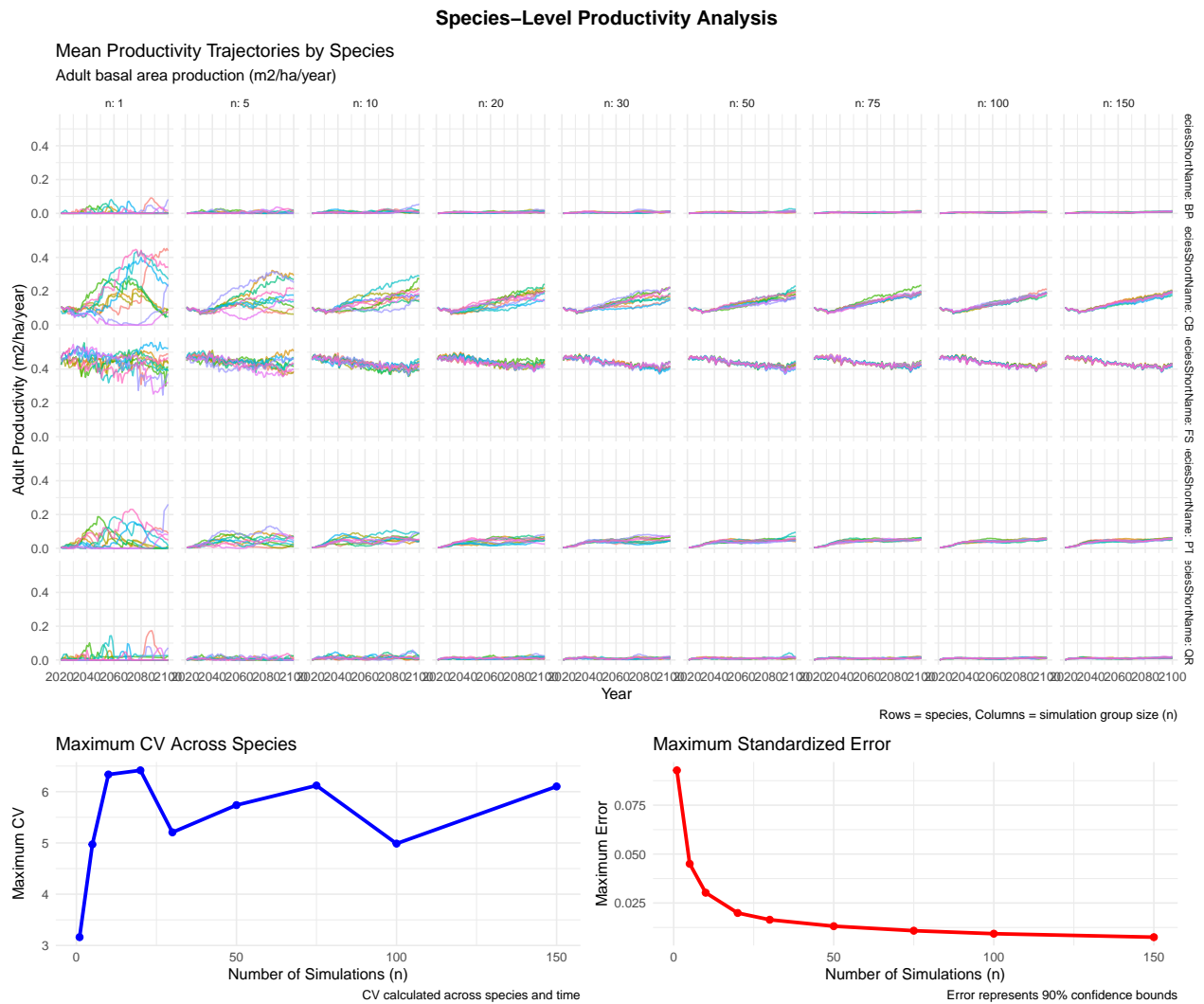


Figure 11: Figure 11: Species-level productivity analysis showing coefficient of variation and error patterns across different simulation group sizes for individual tree species.

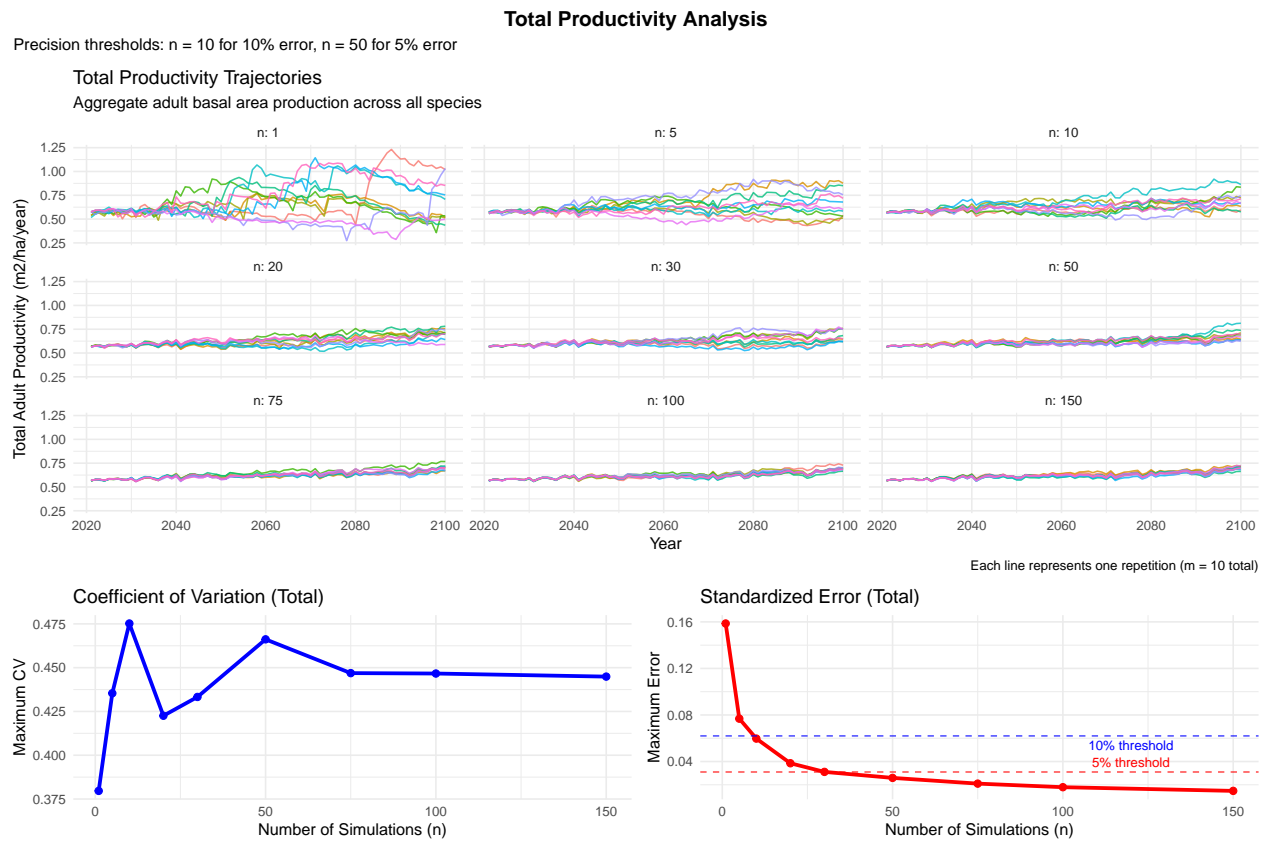


Figure 12: Figure 12: Total productivity analysis across all species, showing convergence patterns for aggregated productivity metrics with precision thresholds at 5% and 10% error levels.

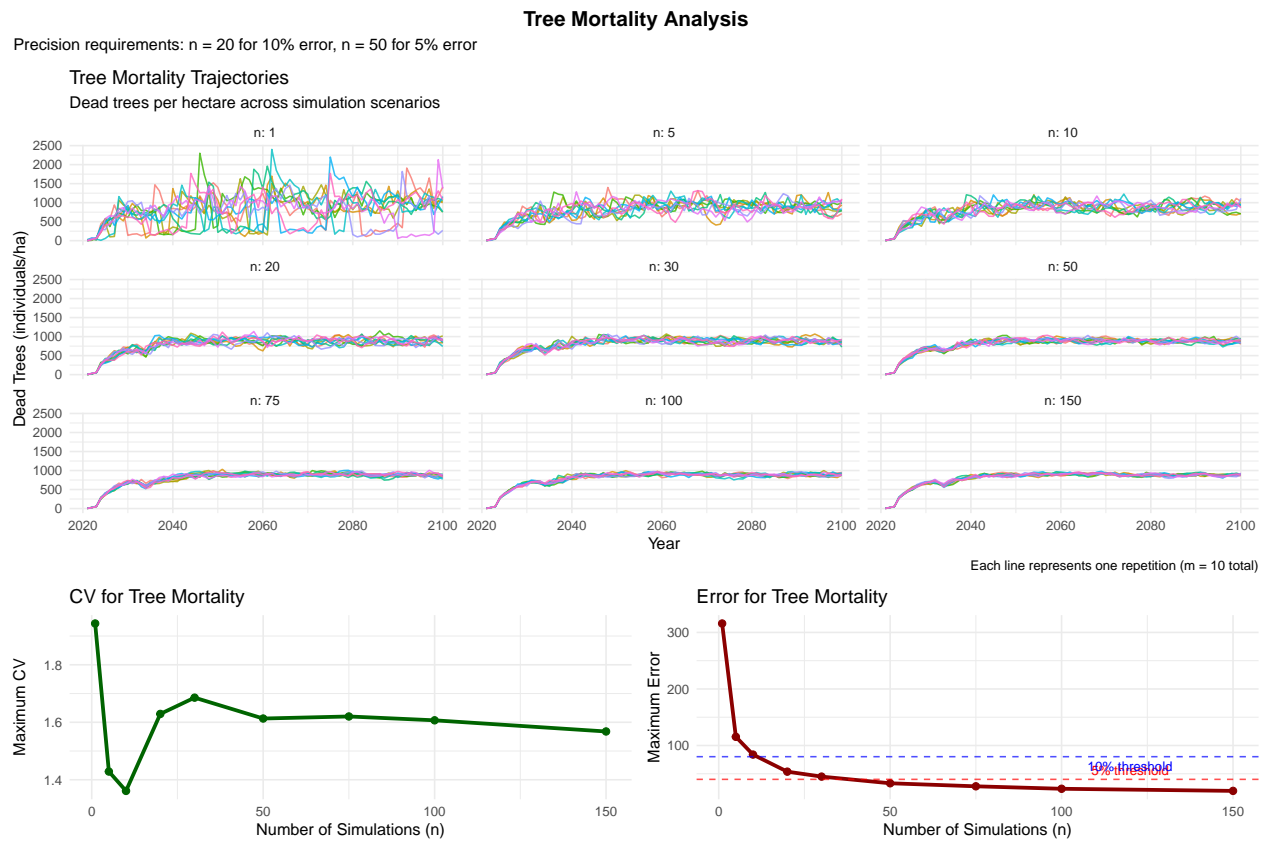
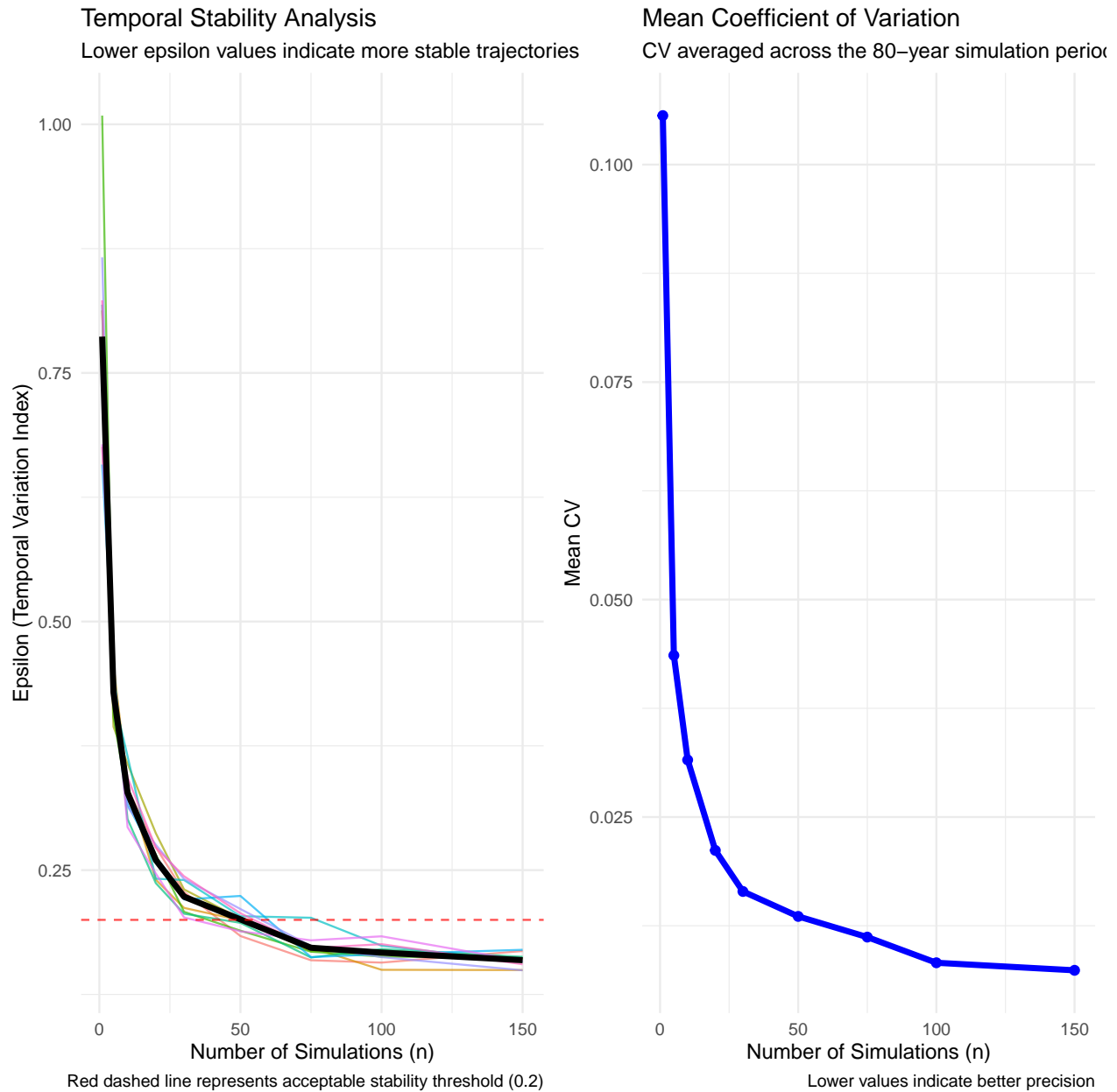


Figure 13: Figure 13: Tree mortality analysis showing the number of simulations required to achieve stable estimates of forest mortality rates with defined precision thresholds.



Discussion

Our analysis supports different simulation parameters based on application requirements:

- 10% error tolerance : $n = 10$, $m = 10$
- 5% error tolerance : $n = 50$, $m = 10$

Methodological Considerations

Strengths

1. **Comprehensive scope:** Analysis covered multiple forest metrics across 80-year simulations
2. **Large sample size:** 5,000 independent simulations provide robust statistical foundation
3. **Systematic approach:** Two-stage optimization ensures both repetition and simulation parameters are optimized

4. **Practical focus:** Results directly applicable to operational forest modeling

Limitations

1. **Single forest type:** Results based solely on Retz forest temperate deciduous ecosystem
2. **Fixed time horizon:** 80-year simulation period may not capture long-term dynamics
3. **Limited climate scenarios:** Analysis used current climate conditions only