

# STOCHASI

Stochastic Chronological Artifact Simulation

Version 1.0

## Technical Documentation

*Mathematical Foundations, Methodology, and Application Guide*

A Streamlit-based Tool for Archaeological Research

January 2026

# Table of Contents

<b>1. Introduction.....</b>	<b>3</b>
1.1 Overview.....	3
1.2 Purpose and Applications.....	3
1.3 Supported Artifact Types.....	3
<b>2. Mathematical Foundations.....</b>	<b>4</b>
2.1 The Replacement Model.....	4
2.1.1 Basic Formulation.....	4
2.1.2 Physical Interpretation.....	4
2.2 Monte Carlo Simulation.....	5
2.2.1 Stochastic Perturbation.....	5
2.2.2 Statistical Aggregation.....	5
2.3 Market Data Interpolation.....	5
<b>3. Simulation Algorithm.....</b>	<b>6</b>
3.1 Algorithm Overview.....	6
3.1.1 Single Run Algorithm.....	6
3.2 Settlement Mode.....	6
3.2.1 Classic Mode.....	6
3.2.2 New Settlement Mode.....	7
3.3 Normalization.....	7
<b>4. Application Structure.....</b>	<b>8</b>
4.1 Data Input Requirements.....	8
4.1.1 Market Supply Data.....	8
4.1.2 Excavation Spectrum Data.....	8
4.2 Simulation Parameters.....	9
4.3 Visualization Outputs.....	9
<b>5. Statistical Analysis Methods.....</b>	<b>10</b>
5.1 Comparison Metrics.....	10
5.1.1 Mean Absolute Error (MAE).....	10
5.1.2 Root Mean Square Error (RMSE).....	10
5.1.3 Maximum Deviation.....	10
5.2 Uncertainty Quantification.....	10
<b>6. Data Export and Persistence.....</b>	<b>11</b>
6.1 Export Formats.....	11
6.1.1 CSV Export.....	11
6.1.2 JSON Configuration.....	11

6.1.3 Graphics Export.....	11
<b>7. Technical Reference.....</b>	<b>12</b>
7.1 Software Dependencies.....	12
7.2 Performance Considerations.....	12
7.3 Reproducibility.....	12
<b>8. Appendix: Terra Sigillata Categories.....</b>	<b>13</b>

# 1. Introduction

## 1.1 Overview

STOCHASI (**S**tochastic **C**hronological **A**rtifact **S**imulation) is a sophisticated computational tool designed for archaeological research. It enables researchers to model and simulate the temporal transformation of artifact spectra, particularly focusing on ceramic assemblages such as Terra Sigillata. The application employs stochastic modeling techniques combined with Monte Carlo simulation methods to predict how artifact distributions evolve over time under various market and cultural conditions.

The fundamental premise of STOCHASI is that archaeological assemblages are dynamic entities. The composition of artifacts at any settlement changes over time due to continuous processes of acquisition, use, breakage, and replacement. By modeling these processes mathematically, STOCHASI allows archaeologists to generate expected artifact distributions for specific time periods, compare these predictions with actual excavation data, and thereby estimate settlement chronologies with quantified uncertainties.

## 1.2 Purpose and Applications

The primary applications of STOCHASI include:

- **Chronological Modeling:** Simulating how artifact spectra evolve from a given starting point, accounting for market supply dynamics and replacement processes.
- **Settlement Dating:** Comparing simulated artifact distributions with excavation data to estimate the most likely occupation period of archaeological sites.
- **Uncertainty Quantification:** Providing confidence intervals for predictions through Monte Carlo simulation, acknowledging the inherent stochasticity in archaeological processes.
- **Hypothesis Testing:** Evaluating whether observed artifact assemblages are consistent with hypothesized settlement histories.
- **Sensitivity Analysis:** Exploring how different parameters (replacement rates, market conditions) affect expected outcomes.

## 1.3 Supported Artifact Types

While STOCHASI was originally developed for Terra Sigillata analysis (with predefined categories IT, LG, BA, MG, RZ representing different production centers), the application supports arbitrary artifact categories. Users can work with:

- Terra Sigillata (Italian, La Graufesenque, Banassac, Central Gaulish, Rheinzabern)
- Any ceramic type classifications
- Small finds categories
- Coin series

- Any other artifact classification system with temporal market supply data

## 2. Mathematical Foundations

### 2.1 The Replacement Model

The core of STOCHASI is a discrete-time replacement model that describes how the artifact composition at a site changes from year to year. The model assumes that in each time step, a fraction of the existing inventory is replaced by new acquisitions drawn from the contemporary market supply.

#### 2.1.1 Basic Formulation

Let  $S(t) = [s_1(t), s_2(t), \dots, s_n(t)]$  be the artifact spectrum at time  $t$ , where  $s_i(t)$  represents the proportion of category  $i$  in the assemblage. The spectrum evolves according to:

$$S(t+1) = (1 - r) \cdot S(t) + r \cdot M(t)$$

Where:

- $r$  is the replacement rate ( $0 \leq r \leq 0.5$ ), representing the fraction of artifacts replaced per year
- $M(t) = [m_1(t), m_2(t), \dots, m_n(t)]$  is the market supply vector at time  $t$
- The sum constraint  $\sum s_i(t) = 100\%$  is maintained through normalization after each step

#### 2.1.2 Physical Interpretation

The replacement rate  $r$  captures the combined effects of artifact breakage, loss, and acquisition of new items. A replacement rate of 10% ( $r = 0.1$ ) implies that approximately one-tenth of the artifact inventory is replaced annually. This parameter integrates multiple archaeological processes:

- **Breakage and loss:** Ceramics break during use, and artifacts may be lost or discarded
- **Natural replacement:** Worn items are replaced with new purchases
- **Consumption patterns:** Household demand for new pottery as family circumstances change
- **Trade activity:** Access to and engagement with regional markets

### 2.2 Monte Carlo Simulation

To account for the inherent randomness in archaeological processes, STOCHASI employs Monte Carlo simulation. Rather than computing a single deterministic trajectory, the model runs multiple simulation iterations with stochastic perturbations.

#### 2.2.1 Stochastic Perturbation

In each simulation step, Gaussian noise is added to the exchanged spectrum:

$$S'(t+1) = S(t+1) + \varepsilon$$

Where  $\varepsilon \sim N(0, \sigma^2)$  is normally distributed noise with standard deviation  $\sigma$  (the **noise parameter**). After adding noise, values are clipped to be non-negative and the spectrum is renormalized to sum to 100%.

## 2.2.2 Statistical Aggregation

After running  $N$  simulation iterations (default: 100, maximum: 500), the results are aggregated to produce:

- **Mean spectrum:** The arithmetic mean across all runs for each time point
- **10th percentile (P10):** Lower bound of the 80% confidence interval
- **90th percentile (P90):** Upper bound of the 80% confidence interval

The 80% confidence band [P10, P90] provides a robust measure of uncertainty, showing the range within which 80% of simulation outcomes fall.

## 2.3 Market Data Interpolation

Market supply data is typically provided at discrete intervals (e.g., decade intervals). STOCHASI uses linear interpolation to estimate market supply values for intervening years:

$$M(t) = M(t_1) + (t - t_1)/(t_2 - t_1) \cdot (M(t_2) - M(t_1))$$

Where  $t_1$  and  $t_2$  are the nearest data points bracketing time  $t$ . This ensures smooth transitions in market conditions and prevents artificial discontinuities in the simulation.

## 3. Simulation Algorithm

### 3.1 Algorithm Overview

The simulation algorithm in STOCHASI processes each Monte Carlo run independently and then aggregates results. The core simulation function implements a vectorized approach for computational efficiency.

#### 3.1.1 Single Run Algorithm

For a single simulation run, the algorithm proceeds as follows:

1. **Initialization:** Set the current spectrum to the initial distribution  $S(0)$
2. **Time Loop:** For each year  $t$  from start year to end year:
  - Retrieve the market supply vector  $M(t)$  from interpolated data
  - Calculate the exchanged spectrum:  $S' = (1-r) \cdot S + r \cdot M$
  - Add Gaussian noise:  $S'' = S' + N(0, \sigma^2)$
  - Clip negative values:  $S''' = \max(S'', 0)$
  - Normalize:  $S(t+1) = S''' \cdot 100 / \text{sum}(S''')$
3. **Storage:** Record the spectrum for each year in the results array

### 3.2 Settlement Mode

STOCHASI offers two operational modes that differ in how the initial year is treated:

#### 3.2.1 Classic Mode

In classic mode, the initial distribution represents the state at the end of the year before the simulation starts. Exchange processes begin immediately from the first year of the simulation period.

#### 3.2.2 New Settlement Mode

In settlement mode, the initial distribution is interpreted as the founding inventory of a new settlement. The first year preserves this initial distribution exactly, with exchange processes beginning only from year 2 onward. This mode is appropriate when modeling sites with a known foundation date.

### 3.3 Normalization

A critical aspect of the simulation is maintaining the constraint that all category proportions sum to 100%. STOCHASI provides two normalization approaches:

- **Automatic normalization:** Applied after every simulation step to ensure mathematical consistency
- **Manual normalization:** Available in the market data editor, allowing users to control when normalization occurs during data editing

The normalization function is defined as:

$$\text{normalize}(x) = x \cdot 100 / \sum x_i$$

## 4. Application Structure

### 4.1 Data Input Requirements

#### 4.1.1 Market Supply Data

Market supply data represents the relative availability of different artifact categories on the regional market over time. This data must be provided as an Excel file (.xlsx or .xls) with the following structure:

Column	Description	Example Values
Year	Time point in years AD/BC	50, 60, 70, ...
Category1	Proportion of first category	0.80, 0.70, 0.55, ...
Category2	Proportion of second category	0.15, 0.25, 0.35, ...
...	Additional categories	...

Data values can be provided as:

- **Proportions (0-1):** Values between 0 and 1, automatically scaled to percentages
- **Percentages (0-100):** Values already expressed as percentages
- **Raw counts:** Arbitrary positive values that will be normalized row-wise

#### 4.1.2 Excavation Spectrum Data

Excavation data represents the actual artifact counts or proportions found at an archaeological site. This data is optional but required for comparison analysis. Supported formats include:

- **Type/Count format:** Two columns with category names and absolute counts
- **Wide format:** Categories as column headers with values in first row

## 4.2 Simulation Parameters

The following parameters control the simulation behavior:

Parameter	Range	Default	Description
Start Year	Variable	125 AD	Beginning of simulation period
End Year	Variable	300 AD	End of simulation period
Replacement Rate	0-50%	10%	Annual fraction of artifacts replaced
Noise (Std. Dev.)	0-10	2.0	Standard deviation of Gaussian noise
Monte Carlo Runs	10-500	100	Number of simulation iterations

Random Seed	0-∞	0	Seed for reproducibility (0 = random)
-------------	-----	---	---------------------------------------

## 4.3 Visualization Outputs

STOCHASI provides several visualization types:

- **Market Supply Plot:** Shows the evolution of market supply proportions over time
- **Simulation Timeline:** Displays mean trajectory with 80% confidence bands
- **Year Spectrum Pie Chart:** Distribution at a selected point in time
- **Comparison Bar Chart:** Side-by-side simulation vs. excavation data
- **Deviation Analysis:** Differences between predicted and observed values

## 5. Statistical Analysis Methods

### 5.1 Comparison Metrics

When comparing simulated spectra with excavation data, STOCHASI calculates several statistical measures to quantify the agreement:

#### 5.1.1 Mean Absolute Error (MAE)

The MAE provides an easily interpretable measure of average prediction error:

$$MAE = (1/n) \cdot \sum |S_i - E_i|$$

Where  $S_i$  is the simulated proportion and  $E_i$  is the excavated proportion for category  $i$ .

#### 5.1.2 Root Mean Square Error (RMSE)

The RMSE gives greater weight to larger errors:

$$RMSE = \sqrt{(1/n) \cdot \sum (S_i - E_i)^2}$$

#### 5.1.3 Maximum Deviation

The category with the largest absolute difference is identified to highlight potential systematic biases:

$$Max\ Dev = max |S_i - E_i|$$

## 5.2 Uncertainty Quantification

The Monte Carlo approach provides natural uncertainty estimates through the distribution of simulation outcomes. The 80% confidence interval [P10, P90] captures the range within which we expect the true value to fall in 80% of cases, given the model assumptions.

Wider confidence bands indicate greater sensitivity to stochastic factors, while narrower bands suggest more deterministic behavior. Categories with wide bands may be more susceptible to random fluctuations in the replacement process.

## 6. Data Export and Persistence

### 6.1 Export Formats

STOCHASI supports comprehensive data export capabilities:

#### 6.1.1 CSV Export

- **Full export:** Includes mean values, replacement rates, and percentiles (P10, P90) for all categories
- **Simple export:** Contains only year and mean spectrum values

#### 6.1.2 JSON Configuration

The JSON export preserves the complete analysis state including:

- All simulation parameters
- Market supply data (including any edits)
- Excavation spectrum data
- Category definitions and names
- Comparison year setting

#### 6.1.3 Graphics Export

Visualizations can be exported in multiple resolutions:

Preset	Dimensions	Use Case
Screen (72 DPI)	1200 × 800	Web display, presentations
Print (300 DPI)	2400 × 1600	Publications, reports
Print (600 DPI)	3600 × 2400	High-quality prints
Poster (A3/A2)	4800 × 3200	Posters, large displays

Both PNG (raster) and SVG (vector) formats are available for each visualization.

## 7. Technical Reference

### 7.1 Software Dependencies

STOCHASI is built on the following technology stack:

Package	Purpose	Version
Streamlit	Web application framework	$\geq 1.0$
NumPy	Numerical computation	$\geq 1.20$
Pandas	Data manipulation	$\geq 1.3$
Plotly	Interactive visualization	$\geq 5.0$
Kaleido	Static image export	$\geq 0.2$

### 7.2 Performance Considerations

The simulation employs vectorized NumPy operations for efficiency. Key optimizations include:

- **Vectorized computation:** All category updates computed simultaneously
- **Pre-allocated arrays:** Memory allocated once before simulation loops
- **Cached interpolation:** Market data interpolated once and reused
- **Efficient RNG:** NumPy's default\_rng for high-quality random numbers

### 7.3 Reproducibility

For reproducible results, STOCHASI provides a random seed parameter. When set to a non-zero value, each Monte Carlo run uses a deterministic seed (`base_seed + run_number`), ensuring identical results across repeated analyses. A seed of 0 indicates random (non-reproducible) mode.

## 8. Appendix: Terra Sigillata Categories

The default configuration includes five Terra Sigillata production centers, representing major ceramic industries of the Roman period:

Code	Full Name	Production Period	Region
IT	Italian	ca. 50 BC – 100 AD	Italy (Arezzo, Pisa)
LG	La Graufesenque	ca. 20 – 120 AD	Southern Gaul
BA	Banassac	ca. 80 – 180 AD	Southern Gaul
MG	Central Gaulish	ca. 100 – 250 AD	Central Gaul (Lezoux)
RZ	Rheinzabern	ca. 150 – 280 AD	Germania Superior

These production centers represent a chronological sequence, with Italian sigillata dominating early markets and Rheinzabern products becoming prevalent in later periods. The transition between production centers reflects broader economic and political changes in the Roman Empire.

— *End of Documentation* —