Supplement of Clim. Past, 22, 783–796, 2026
https://doi.org/10.5194/cp-22-783-2026-supplement

Climate of the Past
Open Access
EGU

Supplement of

Information loss in palaeoecological data from process and observer error

Quinn Asena et al.

Correspondence to: Quinn Asena (qasena@wisc.edu)

Supplementary Information

Model parameters

The four scenarios (Table S1) cover a range of 'fast' and 'slow' driver conditions. All scenarios have two drivers: a primary driver (different among scenarios) carrying 0.85 of the overall environmental driver effect and a random walk driver (with the same parameterisation across scenarios) downweighted to 0.15 of the total environmental effect (Table S2). In scenarios 1 and 2, the magnitude of change in the primary driver is insufficient to cause a complete species turnover and generalist species survive the shift in extrinsic conditions. The random walk driver may amplify, or dampen, the effects of the primary driver, but in general is favourable to most species and acts to stabilise the system. In scenarios 3 and 4 the primary linear driver eventually reaches values unfavourable to most species and the system experiences an average loss of species rather than compositional turnover. Continued change beyond this threshold would eventually cause all species to become extinct. For details on simulating drivers, and the underlying model, see Asena, Perry, and Wilmshurst (2024).

Table S1: Four different scenarios were simulated representing different driving conditions. Scenario 1 is the example followed in the main text.

Scenarios	Drivers	Description
Scenario 1: non-linear fast forcing	An abrupt extrinsic driver switching between two constant conditions and randomly changing driver. The random driver is weighted to 0.15 of the total effect.	The primary influence on the system is from the abrupt extrinsic driver causing a shift in ecosystem assemblage. The primary driver alone is not sufficient to cause a complete species turnover, with generalist species surviving both driver conditions, although, the random walk driver can exaggerate or dampen the effects of the primary driver.
Scenario 2: non-linear slow forcing	A driver following a logistic curve combined with a randomly changing driver. The random driver carries 0.15 of the total environmental effect.	The magnitude of change in the logistic driver is the same as scenario 1 and generalist species can survive under both conditions. However, the rate of change between conditions is slower, causing a more gradual shift in assemblage. The random driver introduces variability and coupled driver effects.
Scenario 3: linear slow forcing	A linearly increasing driver with a random driver weighted to 0.15 of the effect.	The slow linear driver is the primary influence on species assemblages. The extent of change in the linear driver causes a complete, or near-complete, species turnover by the end of the simulation with a low likelihood of generalist species surviving the entire simulation. Species turnover is expected to be gradual in the majority of model runs, although the random walk or combined driver effects can speed up (or slow down) species change.
Scenario 4: linear fast forcing	A linearly increasing driver, with a faster rate of increase than scenario 3, coupled with a random driver. The random driver is weighted to 0.15 of driver effect.	The fast linear driver has a greater magnitude and rate of change than scenario 3, causing the system to move through assemblages faster. In general, the random walk is favourable to most species with a low likelihood of drifting towards extreme values (ralative to the species parameters). However, drift in the random walk, coupled with the linear driver, can cause periods of rapid change or stability.

Table S2: Input values to simulate driving conditions.

Scenario	Drivers	Parameter values				
Scenario 1: non-linear fast forcing	Abrupt	Start value = 98	End value = 102	Onset of change = 3000	Weight = 0.85	
	Random walk	Mean = 100	Standard deviation = 0.2		Weight = 0.15	
Scenario 2: non-linear slow forcing	Logistic	Start value = 98	End value = 102	Onset of change= 1000	Weight = 0.85	Growth rate = 0.003
	Random walk	Mean = 100	Standard deviation = 0.2		Weight = 0.15	
Scenario 3: linear slow forcing	Linear	Start value = 95	End value = 110	Rate of change per time-step = 0.003	Weight = 0.85	
	Random walk	Mean = 100	Standard deviation = 0.2		Weight = 0.15	
Scenario 4: linear fast forcing	Linear	Start value = 90	End value = 115	Rate of change per time-step = 0.005	Weight = 0.85	
	Random walk	Mean = 100	Standard deviation = 0.2		Weight = 0.15	

Species parameters are the same across the four scenarios (Table S3). For details around simulating pseudoproxies see Asena, Perry, and Wilmshurst (2024).

Table S3: Input values to simulate species.

Parameter	Distribution	Values		Description
Tolerance optima	Gaussian	Mean = 100	Standard deviation = 20	The optima of a species to a driver. Each species has one tolerance per driver that, along with the breadth parameter, make up the species' niche.
Tolerance breadth	Gaussian	Mean = 20	Standard deviation = 1	The breadth of each species' tolerance to each driver is drawn from a Gaussian distribution and defines the how generalist a species is to a given driver.
Carrying capacity	Gaussian	Mean = 4000	Standard deviation = 500	Total possible abundance per species that varies through time. Carrying capacity is truncated at 5 × SD.
Carrying capacity variability	Uniform	Min = 0.8	Max = 1	The proportion that the carrying capacity varies per time-step. Carrying capacity variability is not temporally correlated.
Maximum population growth rate per driver	Uniform	Min = 1.01	Max = 1.10	A maximum population growth rate is calculated per driver (i.e., the maximum possible growth rate at the peak of the tolerance); the more drivers, the higher the possible product of the growth rates. Growth rates are tuned according to the number of drivers and the taxa they represent.
Dispersal probability		Probability = 0.01		The probability for each species per time-step of an addition to the species' abundance via dispersal from an outside population.
Dispersal size	Uniform	Min = 1	Max = 10	Number of individuals added to a species' abundance via dispersal from an outside population.
Disturbance probability		Probability = 0.002		Probability of a disturbance event occurring each time-step.
Disturbance size	Gamma (exponential)	Shape = 1	Rate = 8	Size of the disturbance is a proportional loss of abundance with a value drawn from a gamma distribution for each species (with larger disturbances becoming exponentially less likely).

Table S4: List of features extracted from principal curves and Fisher's Information time series and their description. The scenario column indicates which of the four scenarios retained the feature after dropping highly correlated features. "F1" indicates that the feature was extracted from the Fisher's Information time-series for scenario 1. "P1" indicates that the feature was extracted from the principal curve for scenario 1.

Metric	Description	Scenario
Maximum	Maximum value in series	
Minimum	Minimum value in series	F4, F3, F2, F1
Range	Range of series	
Mid-range	Mid-range of series	
Mean	Mean of series	F4, P3
Standard deviation	Standard deviation of series	
Standard deviation to mean ratio	Ratio of the standard deviation to the mean	
Variance	Variance of the series	
Skew	Skewness of the series	F4, F2, F1
Kurtosis	Kurtosis of the series	
Maximum percentage difference from median	Largest percentage difference between maximum magnitude and the median	
Minimum percent difference from median	Largest percentage difference between minimum magnitude and the median	
Absolute percent difference from median	Largest percentage difference between either the maximum or min magnitude and the median	
Maximum percentage difference from mean	Largest percentage difference between maximum magnitude and the mean	F1
Minimum percent difference from mean	Largest percentage difference between minimum magnitude and the mean	F4, F3, P4, P3, P2, P1
Absolute percent difference from mean	Largest percentage difference between either the max or min magnitude and the mean	F1
Absolute rate of change	Greatest percent rate of change either positive or negative	
Maximum rate of change	Greatest percent rate of change	F4, F3, F2, F1, P4, P2
Minimum rate of change	Smallest percent rate of change	F4, F3, F2, F1, P2, P1
Longest run of increasing rate of change	Longest period of increasing percent rate of change	F4, F3, F2, F1, P1
Longest run of decreasing rate of change	Longest period of decreasing percent rate of change	F4, F3, F2, F1, P2, P1
Maximum slope	Largest change between two consecutive points	
Minimum slope	Smallest change between two consecutive points	
Maximum absolute slope	Largest change between two consecutive points either positive or negative	
Longest run of increasing slope	Longest period of increasing slope between two consecutive points	
Longest run of decreasing slope	Longest period of decreasing slope between two consecutive points	
Longest run of increasing values	Longest period of monotonic increase	
Longest run of decreasing values	Longest period of monotonic decrease	
Maximum difference between two points in window	Greatest change between two points in moving windows of 10	F4, F3, F2, F1, P3, P2

Minimum difference between two points in window	Smallest change between two points in moving windows of 10	F4, F3, F2, F1, P2, P1
Number of change points in mean and variance	Number of change points identified using the mean and variance and pruned exact linear time (PELT) method	F4, F3, F2, F1, P4, P3, P2, P1
Maximum change point in mean	Largest segment mean using the PELT method	
Minimum change point in mean	Lowest segment mean using the PELT method	F4, F3, F2, F1, P3, P2, P1
Range of change point in mean	Range of segment means	F2, P3, P1
Largest difference between two change points in mean	Greatest change point in mean using the PELT method	F4, F3, F2, F1
Smallest difference between two change points in variance	Smallest change point in mean using the PELT method	F4, F3, F2, F1
Maximum change point in variance	Largest segment variance using the PELT method	
Minimum change point in variance	Smallest segment variance using the PELT method	F4, F3, F2, F1, P2, P1
Range of change point in variance	Range of segment variances	
Largest difference between two change points in variance	Greatest change point in variance using the PELT method	F2, F1, P4
Smallest difference between two change points in variance	Smallest change point in variance using the PELT method	F4, F3, F2, F1, P4
Total number of change points non-parametric	Number of change points identified using non-parametric functions	F4, P4, P2, P1
Longest period of median state	Maximum number of time windows with the same number of median states calculated by Fisher's information	F4, F3, F2, F1
Number of median states in longest period	Median number of states in the longest period of the same median state	F4, F3, F2, F1
Longest period of mean state	Maximum number of time windows with the same number of mean states calculated by Fisher's information	F4, F3, F2, F1
Number of mean states in longest period	mean number of states in the longest period of the same mean state	F4, F3, F2, F1
Range of cumulative sum	Range of the cumulative sum divided by number of observations multiplied by the standard deviation	
Maximum of cumulative sum	Maximum of the cumulative sum divided by number of observations multiplied by the standard deviation	F2
Minimum of cumulative sum	Maximum of the cumulative sum divided by number of observations multiplied by the standard deviation	F4, F3, F1
Von Newmann variance index	The von Neumann variance index. Requires equally spaced measurements	Dropped due to high correlation
Von Newmann variance index time invariant	The von Neumann variance index accounting irregular intervals	P4, P3, P1

Autocorrelation length	Autocorrelation function where its value is smaller than exp(-1)	F2, P2, P1
Median buffer range percentage	Fraction of photometric points within amplitude div 10 of the median magnitude	P4, P3, P2, P1
Beyond 1 standard deviation	Percentage of points beyond one standard deviation from mean	F2, F1, P4, P3, P2, P1
Fraction increase	The fractions of increasing first differences	P1
Fraction decrease	The fractions of decreasing first differences	Dropped due to high correlation
Fraction ratio	Ratio of the fractions of increasing and decreasing first differences	F4, F3, F2, F1, P4, P3, P2
Small kurtosis	Kurtosis for small sample sizes	Dropped due to high correlation
Flux percentile ratios	The ratio of the difference between quantiles: mid 20, mid 35, mid 50, mid 65 and mid 80	F4 (80), F2 (80), P4 (20, 80), P3 (65, 80), P2 (20, 35, 65, 80), P1 (35, 65, 80)
Percent difference flux percentile	Difference between the 95th and 5th percentile over the median	F3, P4, P3, P2, P1
Quartile difference	Difference between the first and third quartile	F4, F3, F2, F1, P4, P3, P2, P1
Median absolute deviation	Median discrepancy of the data from the median data	P1
Total number of features used per scenario		F1 = 25, F2 = 26, F3 = 22, F4 = 25, P1 = 21, P2 = 20, P3 = 14, P4 = 14

Scenario 2-4 results

Results for scenario 1 are included in the main paper. Following, are visualisations of the results from scenarios 2, 3, and 4; or details on methods, please see the main paper.

Scenario 2: Fisher Information

Results for the feature analysis of the Fisher Information (FI) across the 31 replicate synthetic cores. Showing the Euclidean distance from the error-free benchmark as uncertainties increase in severity.

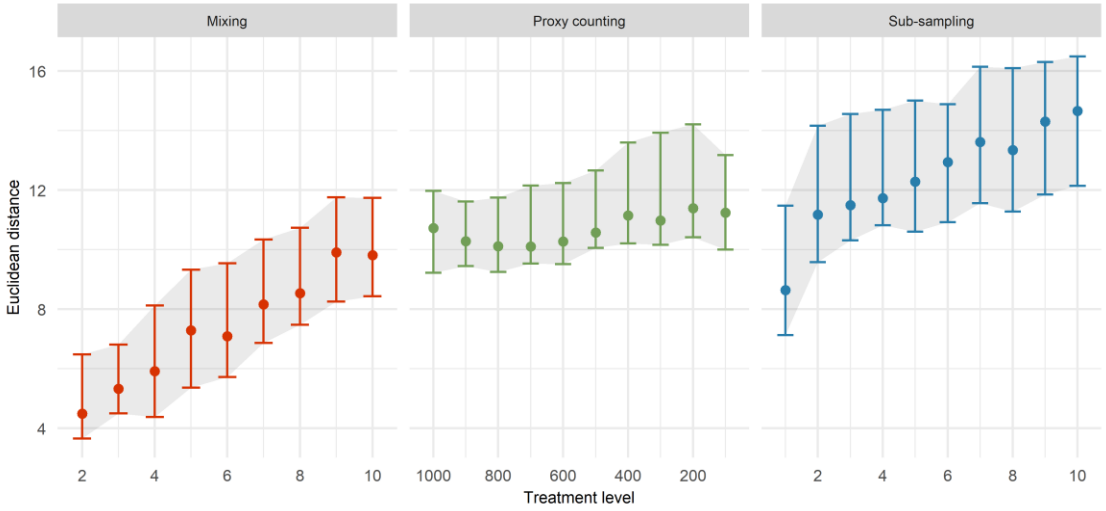

Figure S1: Change in median Euclidean distance on the features extracted from Fisher Information from the 'error-free' core when uncertainties are applied individually.

Results for two combined uncertainties for scenario 2:

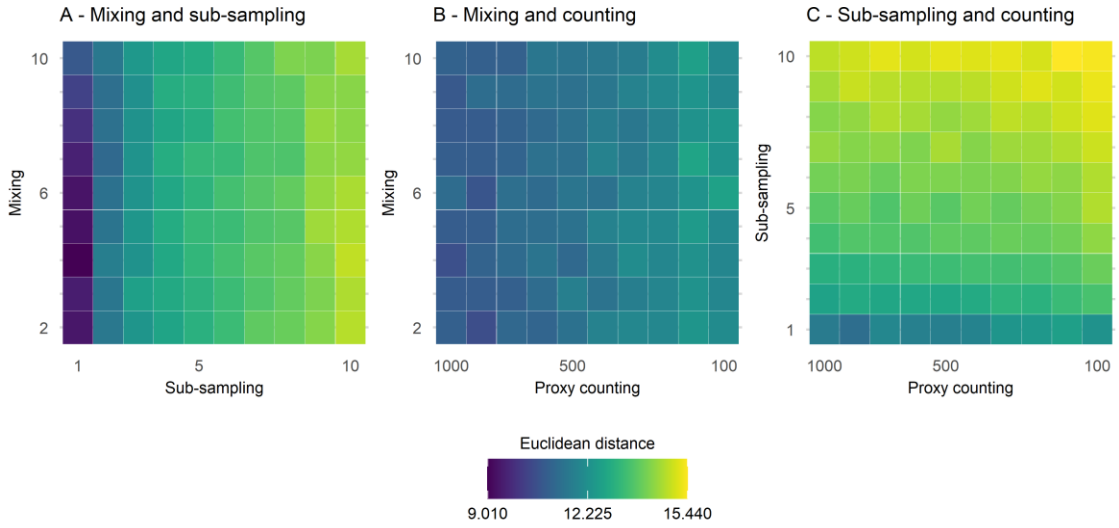

Figure S2: Mean Euclidean distance from the 'error-free' reference core as uncertainty from the treatments increases simultaneously. Mixing in combination with sub-sampling (A) and proxy counting (B). Sub-sampling in combination with proxy counting shows a clear interaction effect as count resolution decreases and sub-sampling intervals increase (C).

When three uncertainties are combined the change in mean Euclidean distance from the simulated counting process is shown across the facets.

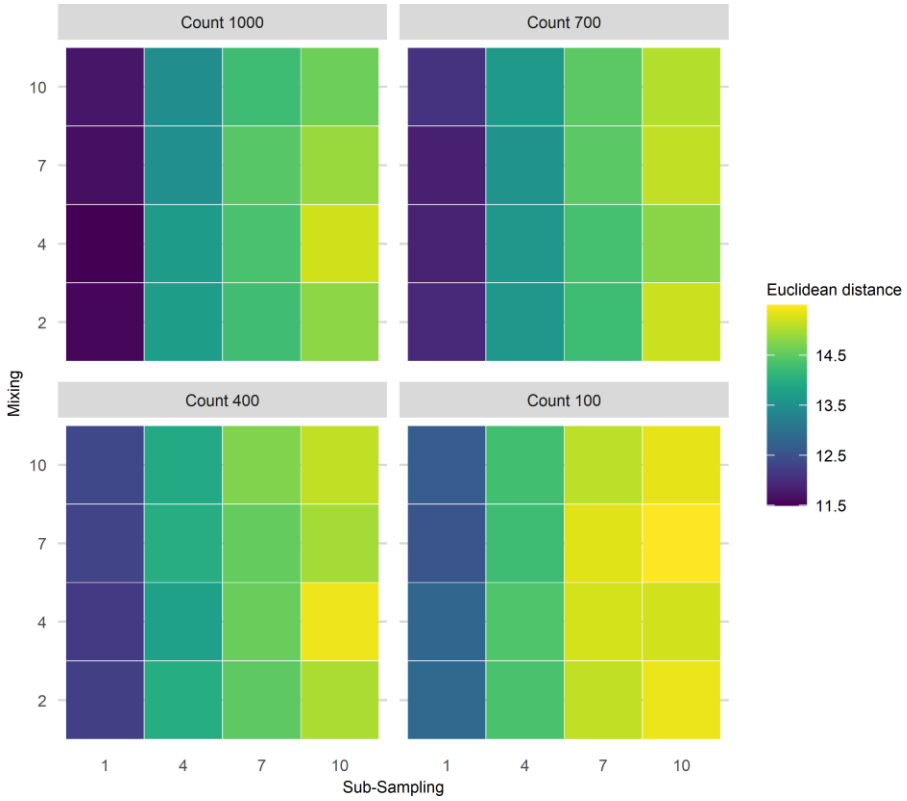

Figure S3: Change in mean Euclidean distance from the 'error-free' core with three treatments applied together. No consistent change in mean Euclidean distance is visible along the mixing axis, indicating no three-way interaction exists among treatments.

Scenario 2: principal curves

The following figures show results from scenario 2 for the feature analysis of the principal curves (PrC) across the 31 replicate synthetic cores.

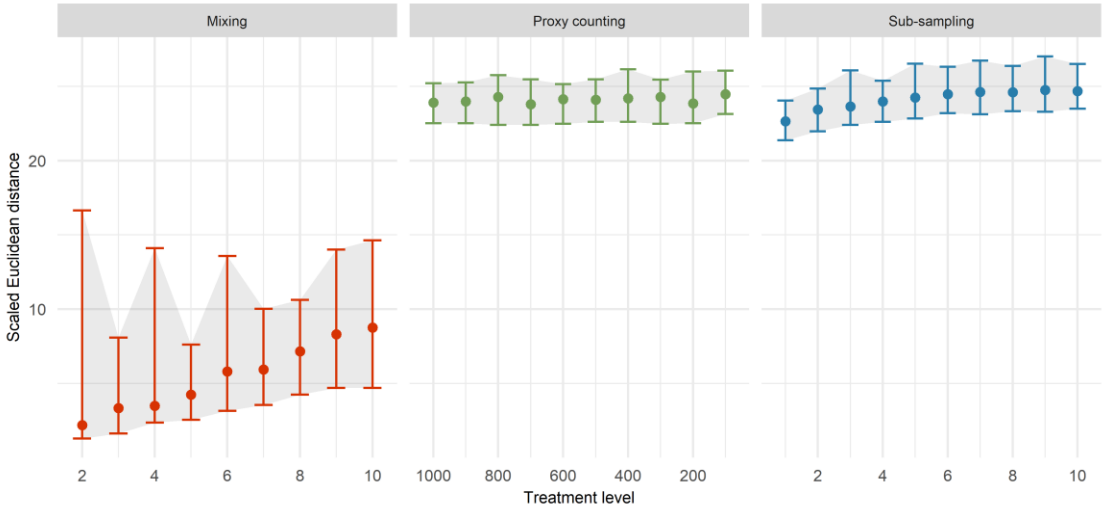

Figure S4: Effect of uncertainties on the median Euclidean distance of the features extracted from the PrCs from the 'error-free' reference.

Applying two uncertainties in combination:

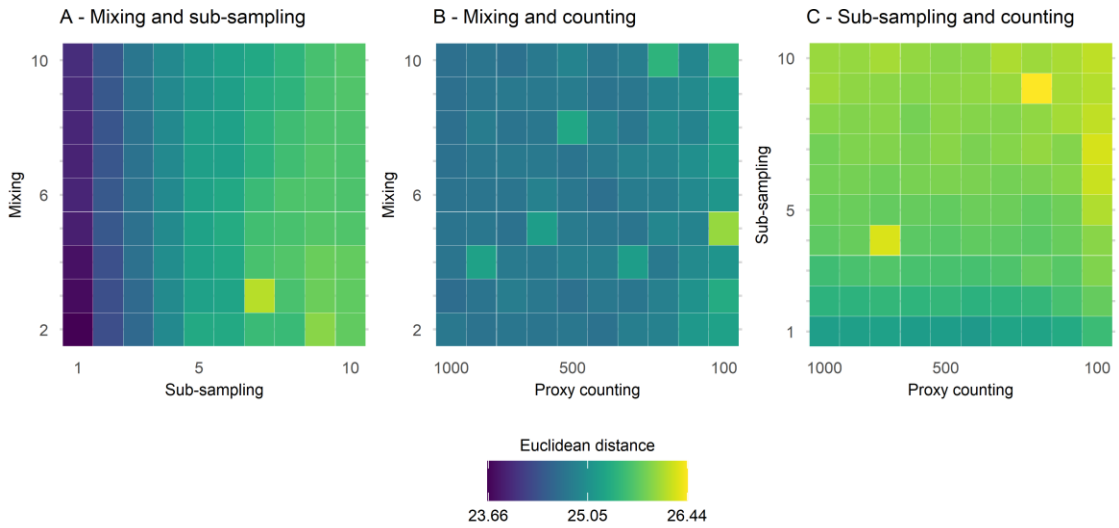

Figure S5: Mean Euclidean distance from the 'error-free' core as treatments are applied in combination to the PrC features.

Finally, the application of all three uncertainties:

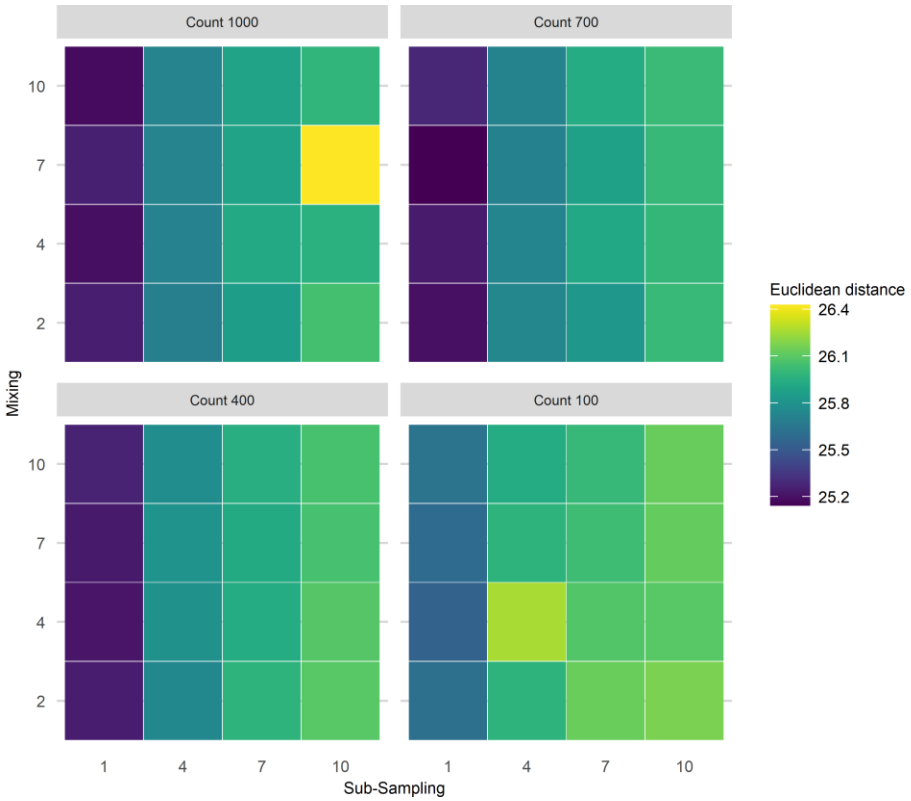

Figure S6: Mean Euclidean distance from the 'error-free' core of all treatments increasing in combination. Proxy count resolution decreases across facets from top left to bottom right. Within each facet the axes show the sub-sampling (frequency in centimeters) and mixing treatments (number of time-steps).

Scenario 3: Fisher Information

Results from scenario 3 (Table S2), showing the change in Euclidean distance from the 'error-free' core as uncertainties are introduced.

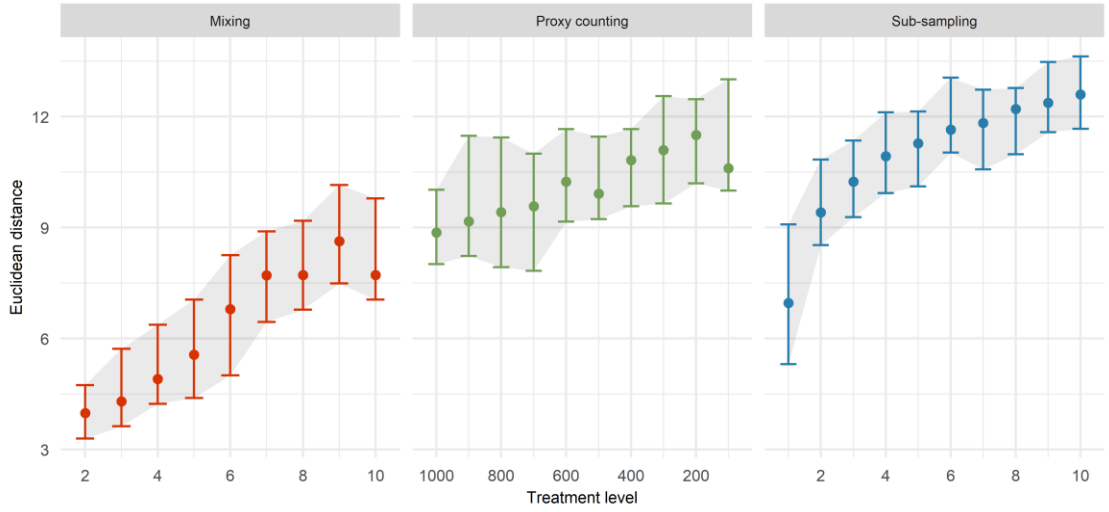

Figure S7: Change in median Euclidean distance from the 'error-free' benchmark in the features extracted from Fisher Information for each individual uncertainty applied at increasing levels.

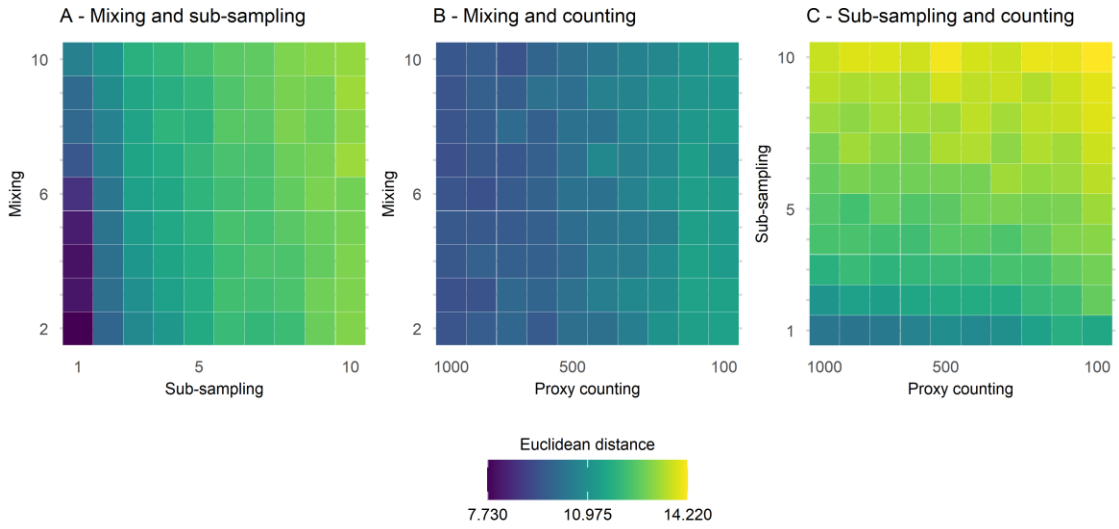

Figure S8: The increase in mean Euclidean distance from the 'error-free' core is shown for each combination of treatments. The weakest effect is observed from mixing in combination with proxy counting (A) when compared with mixing and sub-sampling (A) or sub-sampling in combination with proxy counting (C) which shows the strongest effect.

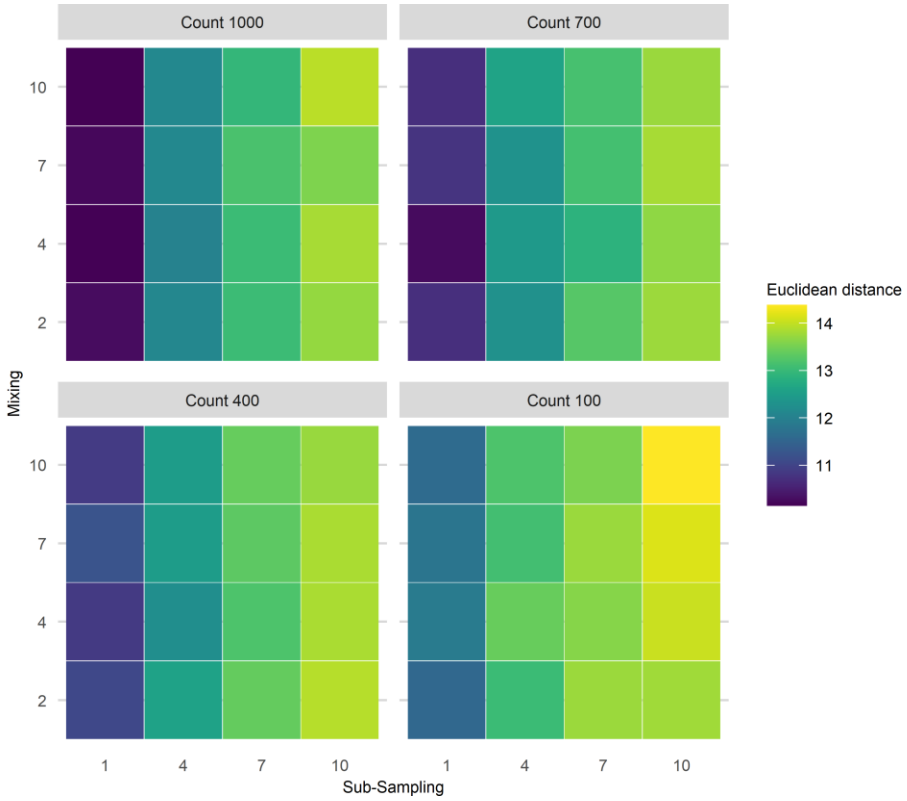

Figure S9: Change in mean Euclidean distance from the 'error-free' core is shown with proxy count resolutions decreasing across plot facets. Each facet shows the effects of mixing and sub-sampling along each axis.

Scenario 3: principal curves

Influence of uncertainties on the extracted features from the principal curves under scenario 3.

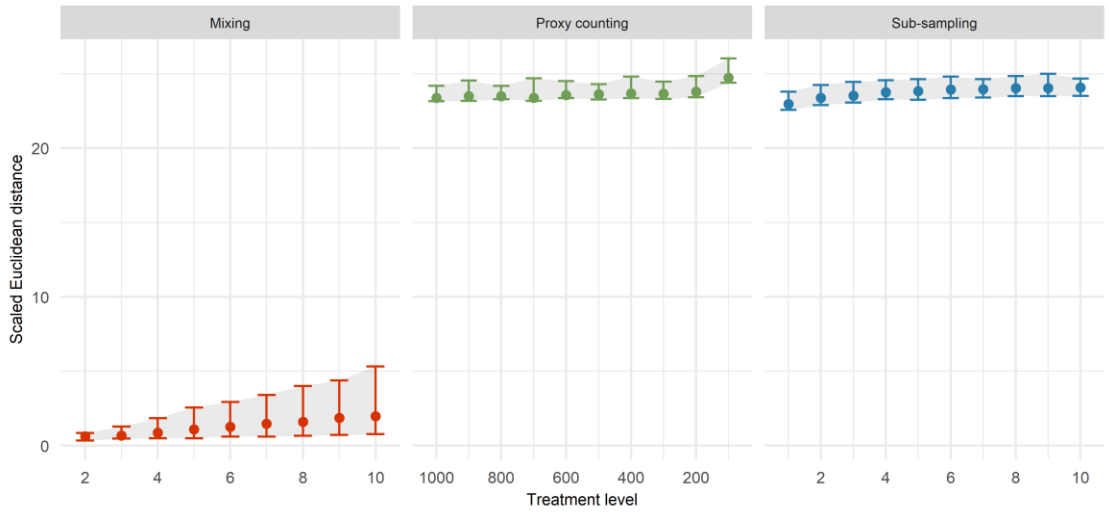

Figure S10: The effect on the extracted features from the principal curves when introducing individual uncertainties to the 'error-free' core. Showing the median, 25th and 75th quantiles.

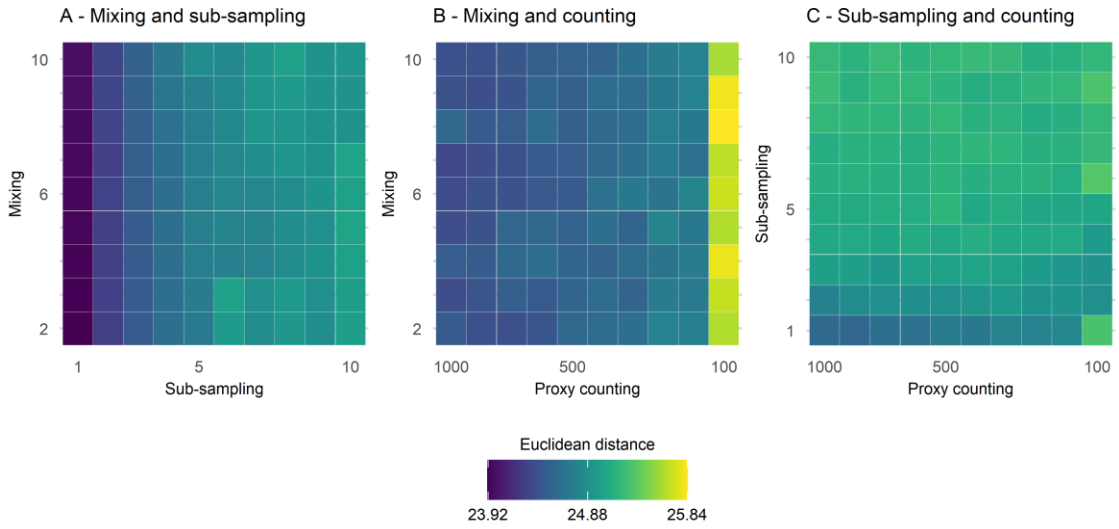

Figure S11: Mean Euclidean distance from the 'error-free' core of PrC features calculated across replicate simulations.

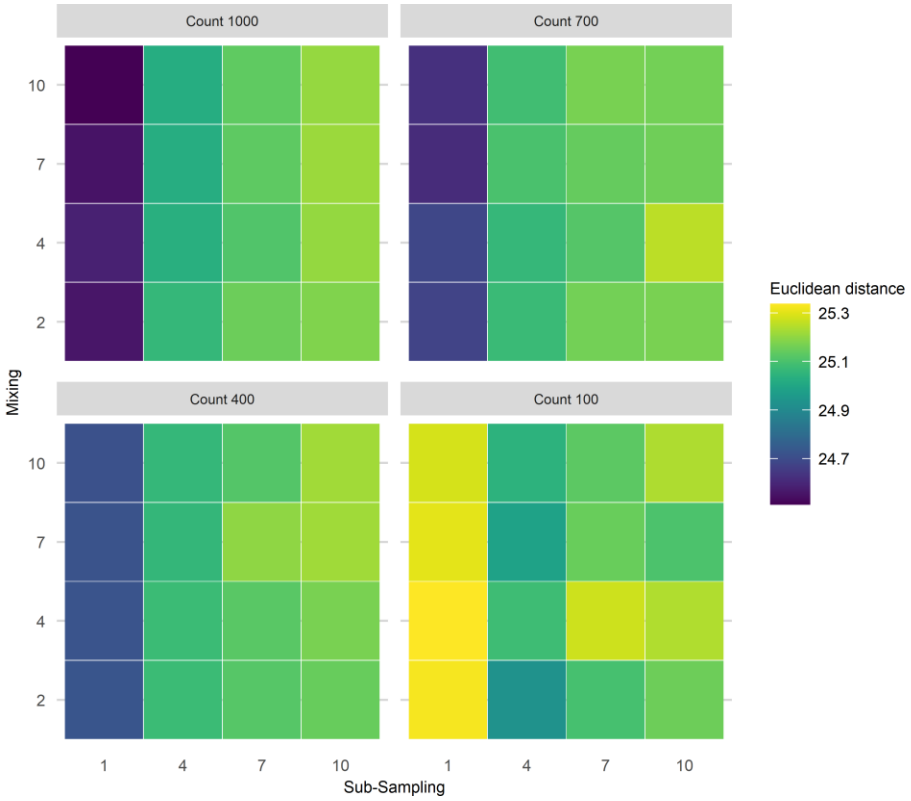

Figure S12: Mean Euclidean distance from the 'error-free' core calculated across replicates from the distances along the PrCs.

Scenario 4: Fisher Information

Results of the extracted features from the Fisher Information series across replicates for scenario 4 (Table S2) as uncertainties are introduced individually, and in combination.

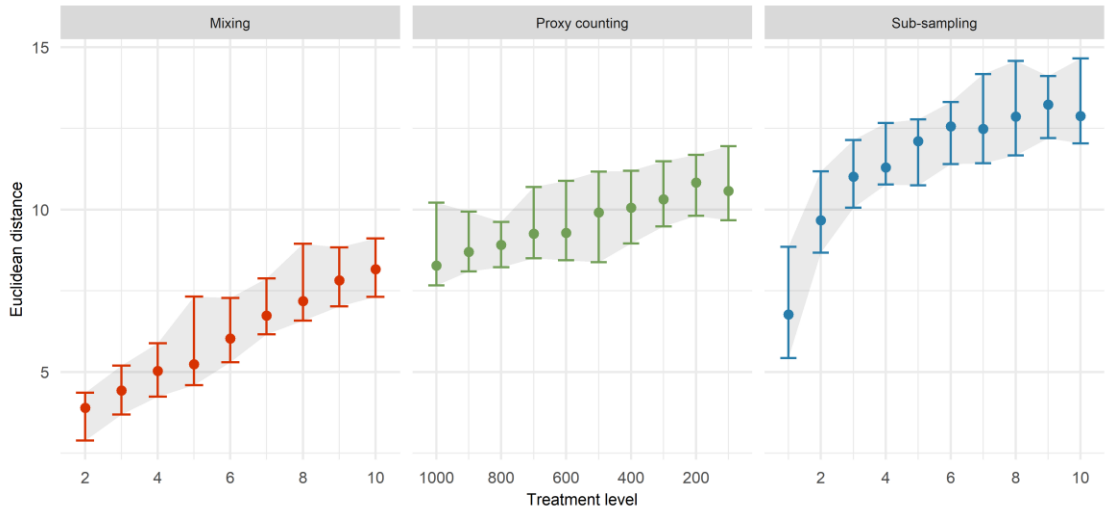

Figure S13: Influence of individual sources of uncertainty on the features extracted from Fisher's information for scenario 4. Showing the median, 25th and 75th quantiles.

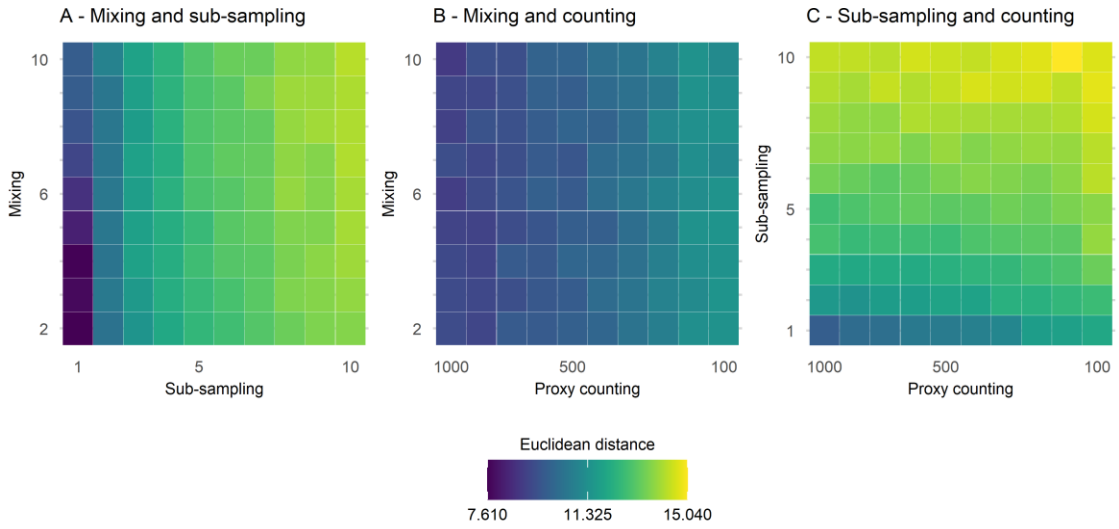

Figure S14: combination. Plots share a single scale to highlight differences among treatment combinations. Sub-sampling in combination with proxy counting (C) shows the greatest total increase in distance, followed by mixing in combination with sub-sampling (A), and finally mixing in combination with proxy counting (B). An interaction effect is clear as sub-sampling and counting treatments increase in severty together (increasing sub-sampling intervals and decreasing proxy count resolution) (C).

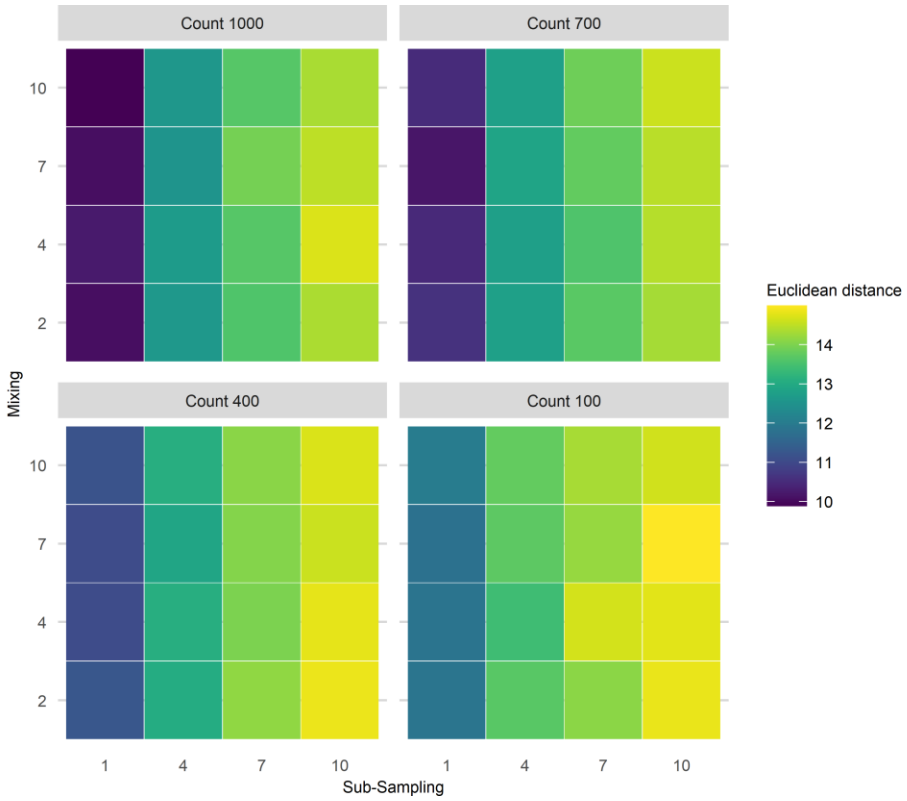

Figure S15: Change in mean Euclidean distance from the 'error-free' core for all three treatments applied simultaneously.

Scenario 4: principal curves

Median Euclidean distance of the extracted features from the principal curves under increasing levels of uncertainty.

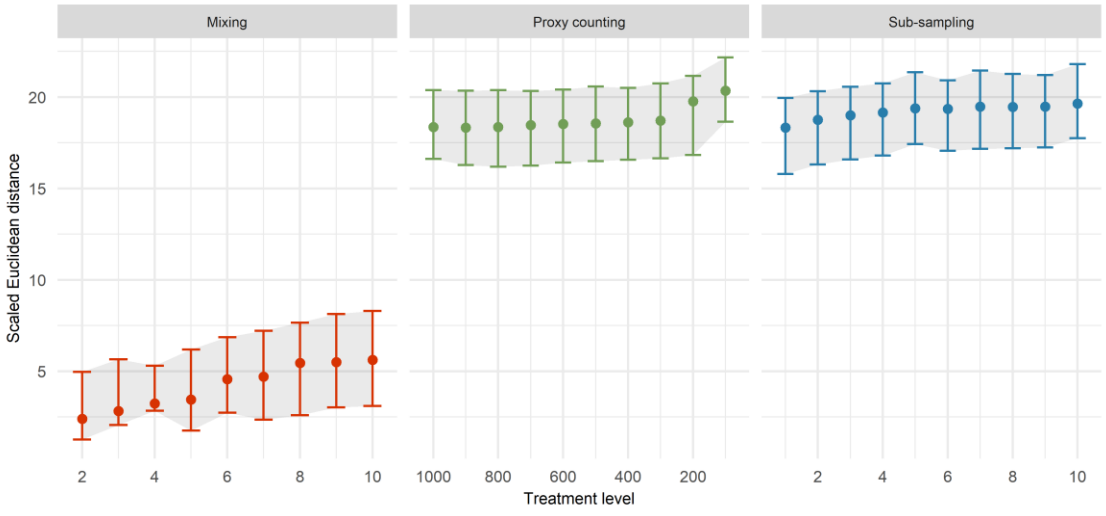

Figure S16: Effect on the mean Euclidean distance on the extracted features from the principal curves as uncertainties are applied individually.

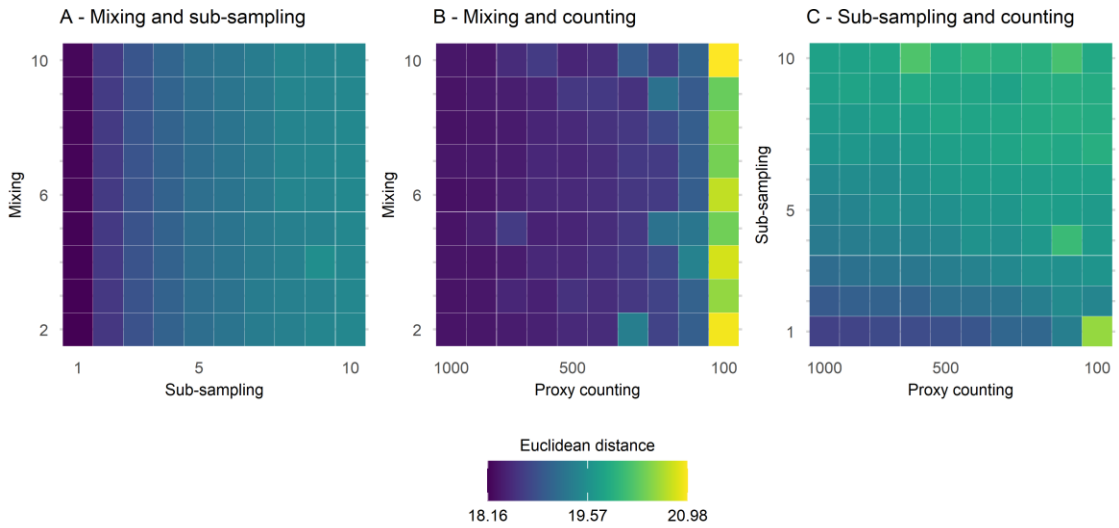

Figure S17: Mean Euclidean distance from the 'error-free' core of the extracted features from the PrCs calculated across replicate simulations.

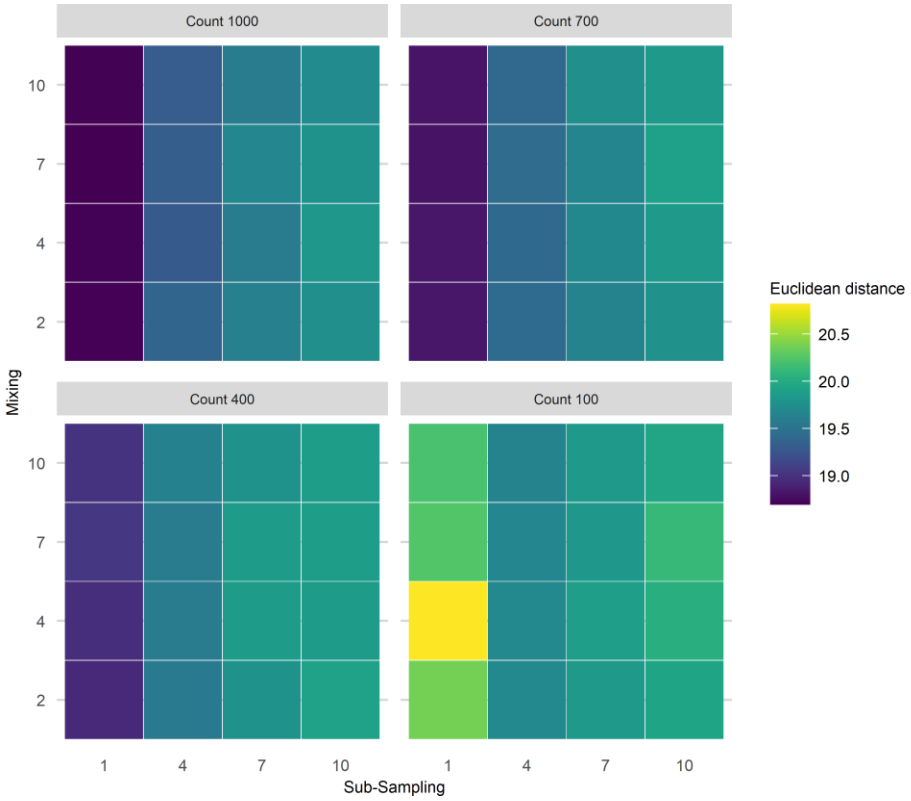

Figure S18: Combined effects of all three treatments increasing in combination shown as the mean Euclidean distance from the 'error-free' core of the extracted features from the distances along the PrCs.

References

Asena, Quinn, George LW Perry, and Janet M Wilmshurst. 2024. "Is the Past Recoverable from the Data? Pseudoproxy Modelling of Uncertainties in Palaeoecological Data." *The Holocene*, May, 09596836241247304. https://doi.org/10.1177/09596836241247304.