

# Validation of the STable AutoCorrelation Integral Estimator (STACIE) Using The AutoCorrelation Integral Drill (ACID) Test Set

model: lorentz()

Gözdenur Toraman,<sup>†</sup> Dieter Fauconnier,<sup>†‡</sup> and Toon Verstraelen<sup>\*¶</sup>

<sup>†</sup> Soete Laboratory, Ghent University, Technologiepark-Zwijnaarde 46, 9052 Ghent, Belgium

<sup>‡</sup> FlandersMake@UGent, Core Lab MIRO, 3001 Leuven, Belgium

<sup>¶</sup> Center for Molecular Modeling (CMM), Ghent University, Technologiepark-Zwijnaarde 46, B-9052,  
Ghent, Belgium

\*E-mail: [toon.verstraelen@ugent.be](mailto:toon.verstraelen@ugent.be)

Version v1.2.1 (2026-05-04 9b7c6de)

## Contents

1. Description of Figures and Tables .....	2
2. Kernel exp1p .....	3
3. Kernel exp1w .....	5
Bibliography .....	7

# 1. Description of Figures and Tables

The following sections contain figures and tables with the same type of results in each section, but computed for different kernels. All figures and tables are labeled with a letter and are explained here. For a full discussion of the results, we refer to the STACIE paper. [1]

## (a) Illustration of input data.

- Left: an example input sequences (first 100 steps).
- Center: the sampling autocorrelation function (ACF) of the input data ( $N = 1024$ ,  $M = 256$ , purple line) and the analytical ACF (dashed line).
- Right: the sampling power spectral density (PSD) of the input data ( $N = 1024$ ,  $M = 256$ , turquoise line) and the analytical PSD (dashed line).

## (b) Scaling of uncertainty of the autocorrelation integral with input data.

- The slope of the slanted gray lines indicates the ideal scaling of the uncertainty (proportional to  $\frac{1}{\sqrt{NM}}$ ). The spacing between the lines corresponds to a factor of 2 in the uncertainty, the ideal case when changing  $N$  by a factor of 4.
- A square represents the standard deviations over 64 repetitions of STACIE's estimate of the autocorrelation integral for a specific combination of  $N$  and  $M$ .
- The dotted lines represent the corresponding predicted uncertainties.

## (c) Assessment of the error estimate of the autocorrelation integral.

- The square blocks show the ratio of the standard deviation of the STACIE estimate and the RMS value of the predicted uncertainty, over 64 repetitions. This value is ideally 100%. Lower values mean that STACIE's predictions have a smaller spread than the predicted uncertainty.
- The dots show the ratio of the mean error and the RMS value of the predicted uncertainty, over 64 repetitions. This value is ideally 0%.

## (d) Scaling of the uncertainty of the exponential correlation time with input data.

- This figure follows the same convention as in (b), but shows results for the uncertainty of the exponential correlation time.

## (e) Assessment of the error estimate of the exponential correlation time.

- This figure follows the same convention as in (c), but shows results for the uncertainty of the exponential correlation time.

## (f) Sensitivity of the autocorrelation integral to the cutoff frequency.

- This plot shows how the autocorrelation integral correlates with the effective number of points used in the fit (top) and the cutoff frequency (bottom).
- Results are shown only for the  $M = 64$ .
- The color code for different  $N$  corresponds to the legends shown in figures (b), (c), (d) and (e).

## (g) Sensitivity of the exponential correlation time to the cutoff frequency.

- The same conventions as in (f) apply, but this figure shows results for the exponential correlation time.

## (h) Number of successful test cases (Failures are typically due to not finding any cutoff frequency with acceptable results.)

### (i) Sanity check counts for the effective number of points

- Number of test cases for each combination of  $N$  and  $M$  where the effective number of points used in the fit is below  $20P = 60$ .

### (j) Sanity check counts for the regression cost z-score

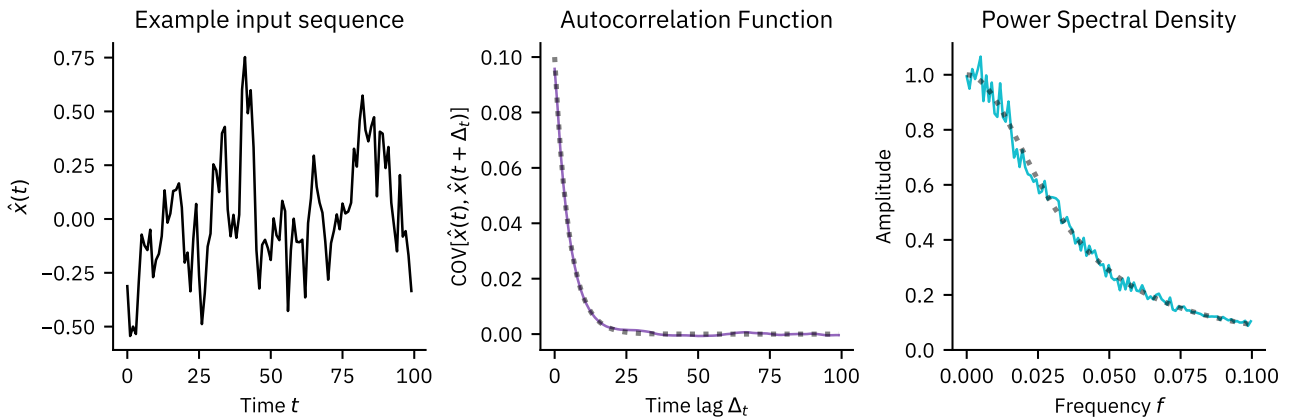
- Number of test cases for each combination of  $N$  and  $M$  where the z-score of the regression cost exceeds 2.

### (k) Sanity check counts for the cutoff criterion z-score

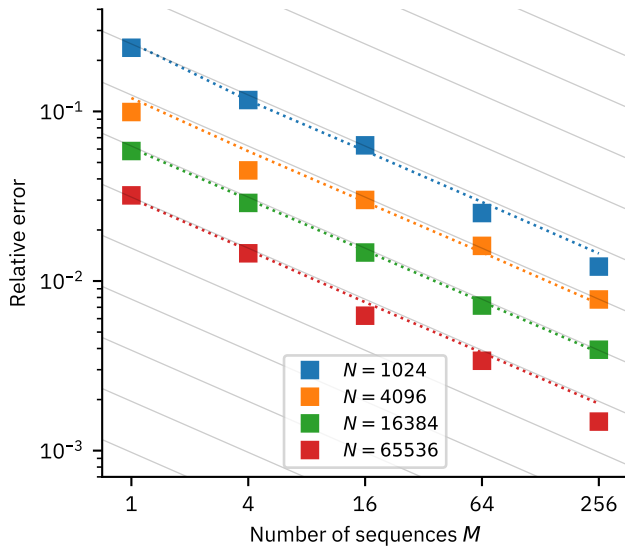
- Number of test cases for each combination of  $N$  and  $M$  where the z-score of the cutoff criterion exceeds 2.

## 2. Kernel exp1p

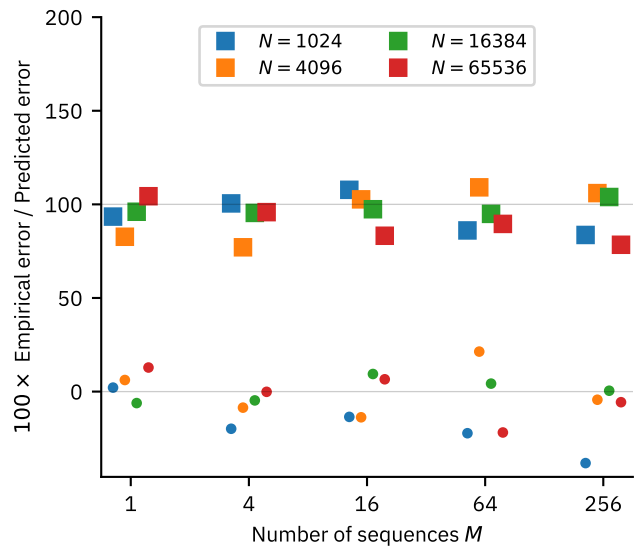
(a) Illustration of input data



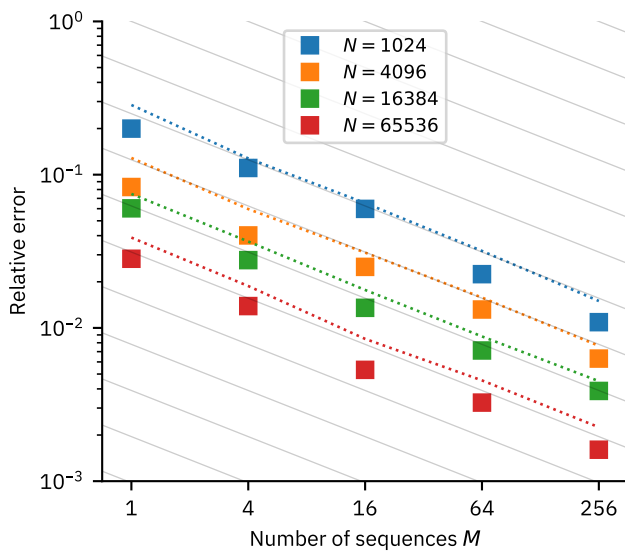
(b) Scaling of uncertainty of the autocorrelation integral with input data



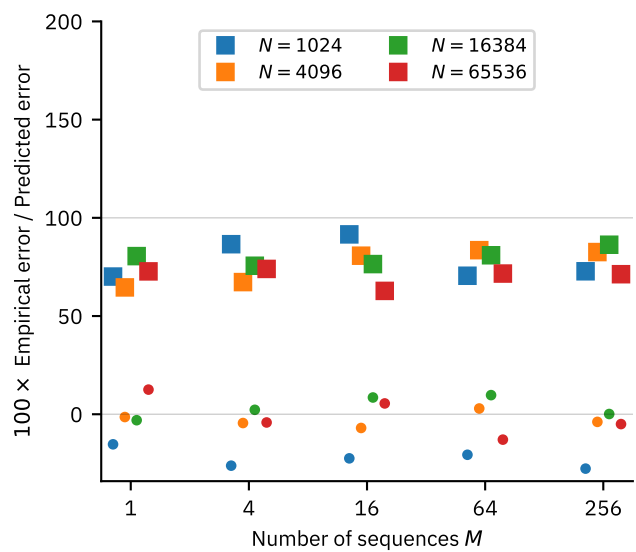
(c) Assessment of the error estimate of the autocorrelation integral



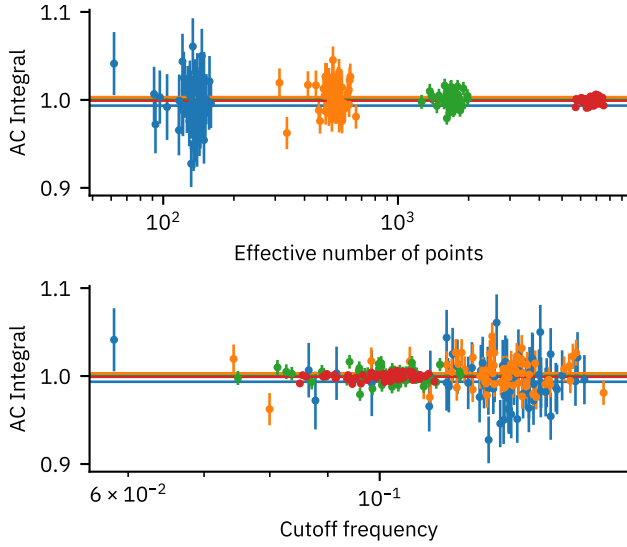
(d) Scaling of uncertainty of the exponential correlation time with input data



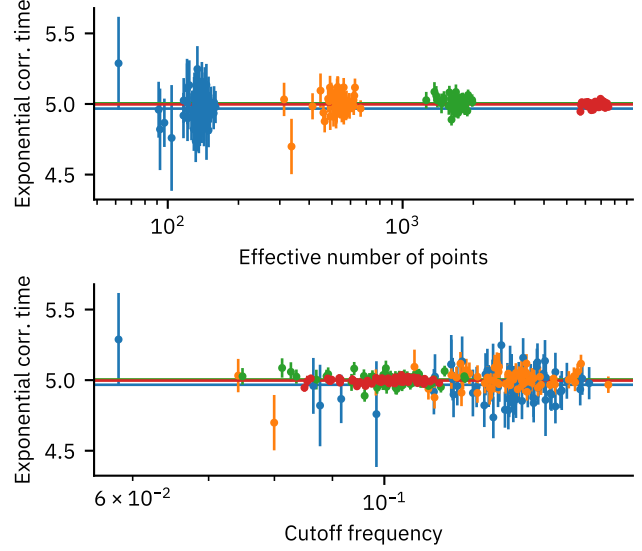
(e) Assessment of the error estimate of the exponential correlation time



**(f) Sensitivity of the autocorrelation integral to the cutoff frequency**



**(g) Sensitivity of the exponential correlation time to the cutoff frequency**



**(h) Number of successful test cases**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	64	64	64	64	64
$N = 4096$	64	64	64	64	64
$N = 16384$	64	64	64	64	64
$N = 65536$	64	64	64	64	64

**(i) Sanity check counts for the effective number of points**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	0	0	0	0	0
$N = 4096$	0	0	0	0	0
$N = 16384$	0	0	0	0	0
$N = 65536$	0	0	0	0	0

**(j) Sanity check counts for the regression cost z-score**

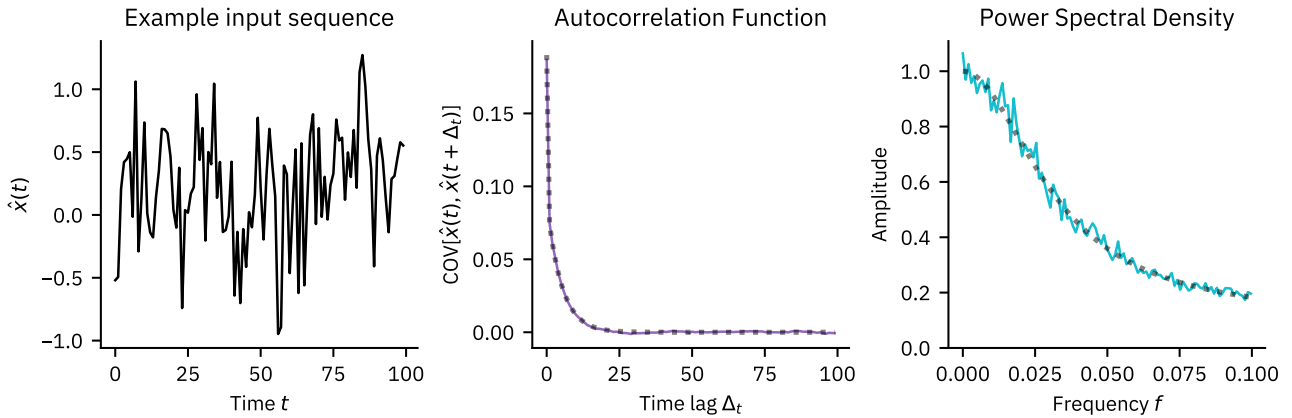
	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	0	0	0	1	1
$N = 4096$	0	1	0	0	0
$N = 16384$	0	0	0	2	1
$N = 65536$	0	0	0	0	2

**(k) Sanity check counts for the cutoff criterion z-score**

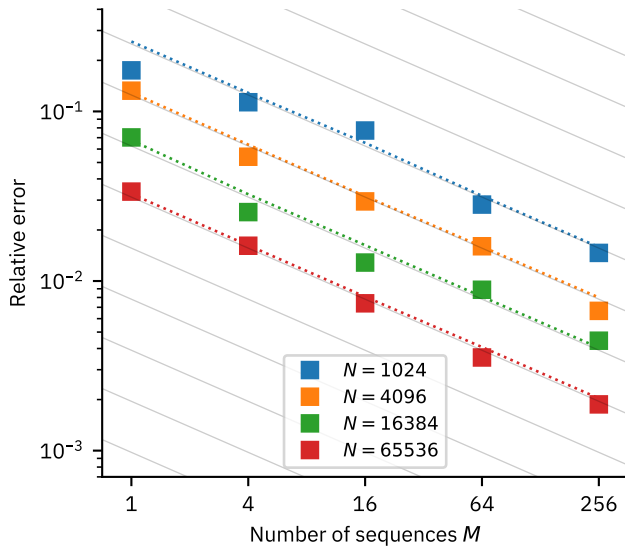
	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	0	0	1	1	0
$N = 4096$	0	0	0	1	1
$N = 16384$	1	0	2	0	0
$N = 65536$	1	1	1	1	0

### 3. Kernel exp1w

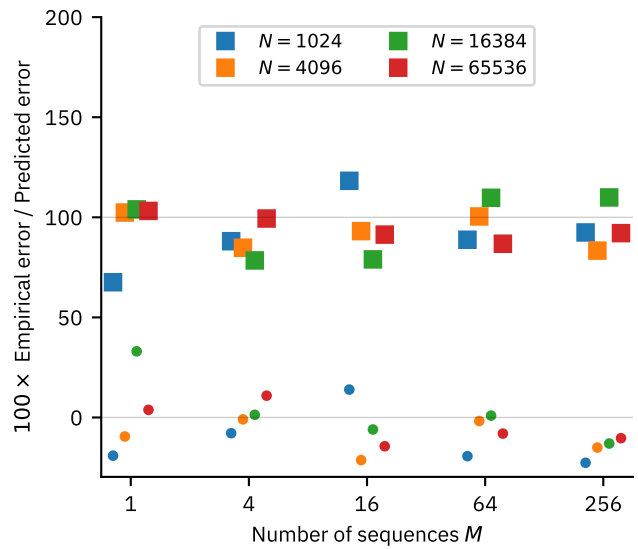
(a) Illustration of input data



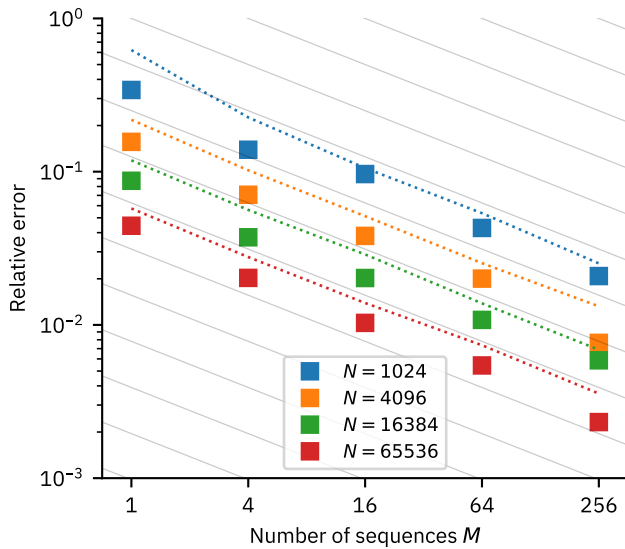
(b) Scaling of uncertainty of the autocorrelation integral with input data



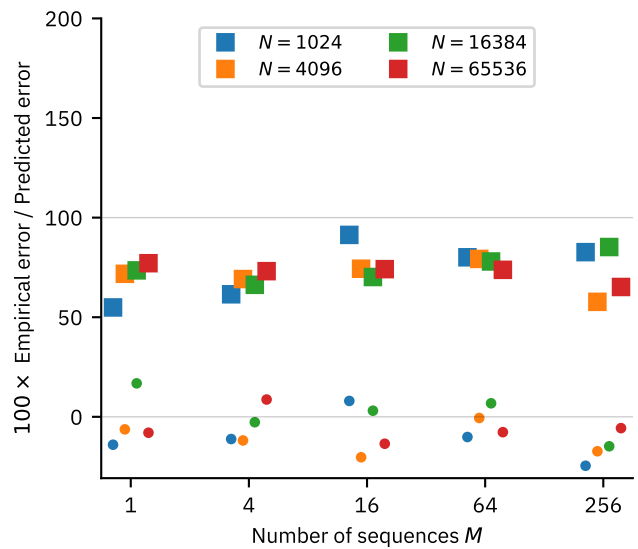
(c) Assessment of the error estimate of the autocorrelation integral



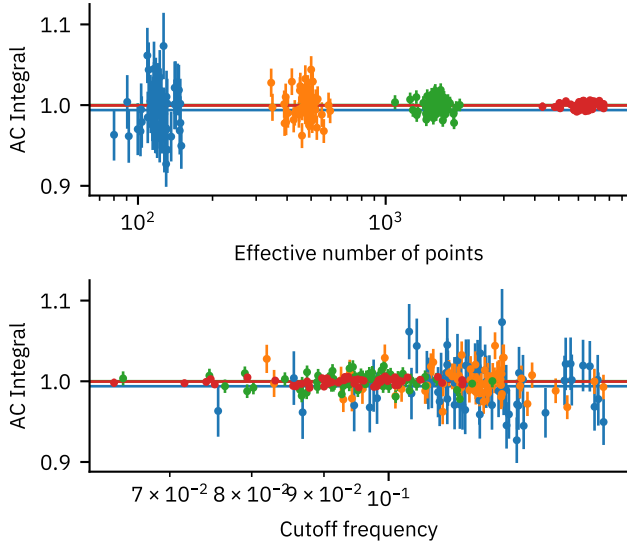
(d) Scaling of uncertainty of the exponential correlation time with input data



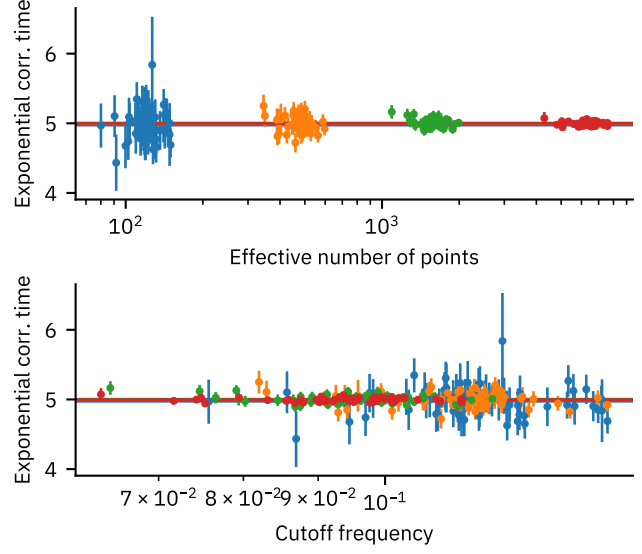
(e) Assessment of the error estimate of the exponential correlation time



**(f) Sensitivity of the autocorrelation integral to the cutoff frequency**



**(g) Sensitivity of the exponential correlation time to the cutoff frequency**



**(h) Number of successful test cases**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	64	64	64	64	64
$N = 4096$	64	64	64	64	64
$N = 16384$	64	64	64	64	64
$N = 65536$	64	64	64	64	64

**(i) Sanity check counts for the effective number of points**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	1	0	0	0	0
$N = 4096$	0	0	0	0	0
$N = 16384$	0	0	0	0	0
$N = 65536$	0	0	0	0	0

**(j) Sanity check counts for the regression cost z-score**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	0	2	1	0	1
$N = 4096$	0	0	2	0	1
$N = 16384$	0	0	0	1	1
$N = 65536$	0	0	0	1	0

**(k) Sanity check counts for the cutoff criterion z-score**

	$M = 1$	$M = 4$	$M = 16$	$M = 64$	$M = 256$
$N = 1024$	0	0	0	1	0
$N = 4096$	0	0	0	0	1
$N = 16384$	0	0	0	0	2
$N = 65536$	1	3	1	0	1

## Bibliography

- [1] G. Toraman, D. Fauconnier, and T. Verstraelen, “STable AutoCorrelation Integral Estimator (STACIE): Robust and accurate transport properties from molecular dynamics simulations,” *Journal of Chemical Information and Modeling*, vol. 65, no. 19, pp. 10445–10464, Sep. 2025, doi: [10.1021/acs.jcim.5c01475](https://doi.org/10.1021/acs.jcim.5c01475).