# Assignment 4: Estimating Errors

Marta Casado Carrasquer

February 26, 2025

## 1 Introduction

In molecular dynamics simulations for physical systems, the initial output usually consists of finite time series of correlated data. When investigating scalar system properties of stationary states, such as the average potential energy, average kinetic energy, or average temperature, the first task in the data analysis is to calculate time averages. Nevertheless, due to their finite duration, they are inclined to fluctuate. Therefore, running the same simulation again will result in a different value for the same parameter. It is for that reason that, the next step in data analysis is to estimate the variance of these finite time averages.

In this report, we go over the so-called "blocking" method where we explain it in Section 2 and analyse its results in Section 3.

## 2 Theory and Methods

### 2.1 Blocking Method

The blocking method is a technique used to reduce autocorrelation in time series data and estimate statistical uncertainties. The idea is to progressively average the data into larger blocks and observe how the standard error converges.

In theory, the method works as follows:

The data is divided into $n_b = N/2^k$ blocks, where $k$ is the blocking level. Each block contains an average of its respective data points:

$$A_i^k = \frac{1}{2}\left(A_{2i-1}^{(k-1)} + A_{2i}^{(k-1)}\right), \quad i = 1, \ldots, n_b \tag{1}$$

where $A_i^{(0)} = A_j$ is the original data.

Furthermore, the variance of the block averages is computed for each blocking level as:

$$\sigma^2(A^k) = \frac{1}{n_b} \sum_{j=1}^{n_b} \left(A_i^k - \bar{A}\right)^2. \tag{2}$$

And the corresponding standard error of $\bar{A}$ is given by:

$$\sigma(\bar{A}) = \frac{\sigma(A^k)}{\sqrt{n_b - 1}} \tag{3}$$

The blocking process continues until statistical convergence is observed, where successive variance estimates become stable, indicating reduced autocorrelation effects.

However, in our implementation, we use a slightly different blocking approach. Instead of recursively defining the block averages, our method increases the block size by doubling it at each step, i.e., $B_k = 2B_{k-1}$. For each block size, we compute the average independently from the original data:

$$A_i^k = \frac{1}{B_k} \sum_{j=(i-1)B_k+1}^{iB_k} A_j, \quad i = 1, \ldots, n_b \tag{4}$$

where $B_k$ is the block size, and $n_b = N/B_k$ is the number of blocks. The variance of the block averages is then computed in the same manner as described previously.

To observe the error bars of the variance in our plots, we approximate Equation (3) for large $N$ and use the following formula:

$$\sigma(\bar{A}) \approx \frac{\sigma(A^k)}{\sqrt{n_b - 1}} \left(1 \pm \frac{1}{\sqrt{2(n_b - 1)}}\right) \tag{5}$$

This implementation allows us to analyse the behaviour of statistical errors as a function of block size and determine the point at which the variance stabilizes.

# 3 Results and Discussion

Here, we present the results obtained from the blocking analysis applied to our Molecular Dynamics (MD) simulation data. The blocking method was used to estimate the standard errors of the average potential energy, kinetic energy, and temperature.

## 3.1 Summary of Results

The simulation was run for 30 ps, i.e. 29999 steps, and equilibrium was reached around 1 ps, i.e. 1000 steps. The calculations for the average energies and temperature as well as the standard errors, are obtained once the steps for equilibrium have been reached.

The average values obtained along with their estimated uncertainties as well as the chosen block-length are as follow:

$\bar{\mathcal{V}} = (-1.18132 \pm 0.00085) \times 10^{-18}$ J

$$\sigma(\bar{\mathcal{V}}) = 0.00085 \times 10^{-18}\,\text{J} = 8.5 \times 10^{-22}\,\text{J}$$

$\bar{\mathcal{K}} = (2.4166 \pm 0.0085) \times 10^{-19}$ J

$$\sigma(\bar{\mathcal{K}}) = 0.0085 \times 10^{-19}\,\text{J} = 8.5 \times 10^{-22}\,\text{J}$$

$\bar{T} = 93.35 \pm 0.33$ K

$$\sigma(\bar{T}) = 0.33\,\text{K}$$

Block Size = 320.

Just for checking:

The expected kinetic energy is given by:

$$\bar{\mathcal{K}}_{\text{expected}} = \frac{3}{2} N k_B \bar{T} \tag{6}$$

Using the reported temperature:

$$\bar{T} = 93.35 \pm 0.33 \text{ K} \tag{7}$$

with $N = 125$ particles and $k_B = 1.380649 \times 10^{-23}$ J/K, we obtain:

$$\begin{aligned} \bar{\mathcal{K}}_{\text{expected}} &= \frac{3}{2} \times 125 \times (1.380649 \times 10^{-23}) \times 93.35 \\ &= 2.4166 \times 10^{-19} \text{ J}. \end{aligned}$$

which matches perfectly with our reported value for $\bar{\mathcal{K}}$.

The corresponding blocking analyses plotted with error bars are:

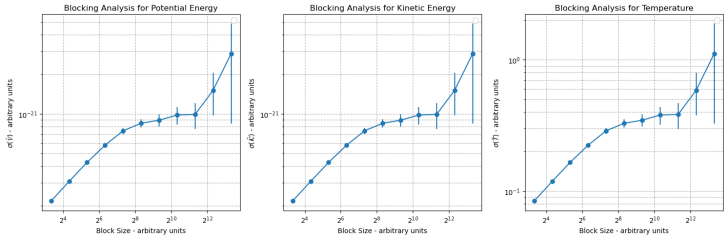


Figure 1: Block Analysis for $\sigma(\bar{\mathcal{V}}), \sigma(\bar{\mathcal{K}})$ and $\sigma(\bar{T})$.

## 3.2 Behaviour of the Blocking Curves

Figure 1 presents the blocking analysis for potential energy, kinetic energy, and temperature. The curves exhibit an initial decrease in standard error (error bars are not visible) as the block size increases. This behaviour indicates that as data points are grouped into larger blocks, short-range correlations are progressively averaged out.

At sufficiently large block sizes, the standard error reaches a plateau (after $2^8$, which can be approximated using the block size = 320). This is a sign of statistical convergence. The plateau suggests that the autocorrelation effects have been effectively reduced and further increases in block size do not significantly impact the estimated uncertainty. However, at very large block sizes, fluctuations appear due to the decreasing number of blocks, which can lead to unreliable variance estimates. This effect is particularly visible in the rightmost regions of the plots, where error bars begin to expand.

## 3.3 Error Bars and Stability of Estimates

The error bars in the plots represent the statistical uncertainties associated with each block size. Initially, they shrink as the block size increases, reflecting improved uncertainty estimation. Once the variance stabilizes, the error bars remain relatively constant.

A critical aspect of blocking analysis is selecting an appropriate block size. If chosen too small, residual correlations may still be present, leading to an underestimation of statistical uncertainty. On the other hand, if the block size is too large, the number of independent blocks becomes insufficient, causing unreliable estimates. In our case, a block size of 320 was selected, ensuring a reasonable balance between statistical independence and variance stability.

## 3.4 Implications for Equilibration

Although the blocking analysis was conducted using data from an already equilibrated simulation, its trends provide insights into the equilibration process. The key indicator of equilibration is the plateau observed in the standard error as a function of block size. If the system were not equilibrated, the standard error would exhibit systematic drifts instead of stabilizing. The presence of a well-defined plateau suggests that the data used for analysis originates from a statistically stationary state.

Thus, blocking analysis confirms that the fluctuations in measured quantities arise from equilibrium statistical variations rather than non-equilibrium transient effects.

Figure 2 from Assignment 3, shows the evolution of the energies in our simulation, before and after equilibrium.

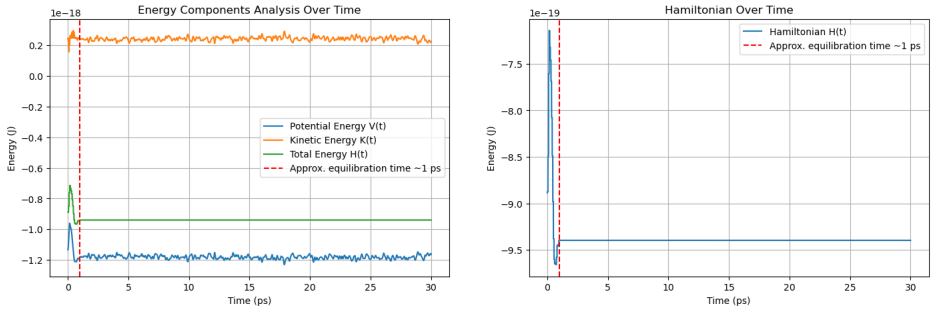


Figure 2: Evolution of $\mathcal{K}, \mathcal{V}$ and $\mathcal{H}$ over time, $t = 30$ps.