2.1 Convergence Analysis of a Time Series Average

Many applications of MD simulation involve the sampling of a system at equilibrium. The quantitative analysis of such simulations typically involves the calculation of averages of properties which can be obtained from individual configurations. These time series averages can be considered estimates of the true ensemble average of the system if: (1) the system is ergodic, (2) the simulation conditions are appropriate e.g. the system boundaries and integration choices, (3) potential energy functions accurately represent the physical system of interest, and (4) the simulation has sampled relevant configurations with the correct probability. Ensemble averages calculated by MD simulation are often compared directly to experimentally properties (e.g. liquid density) or indirectly via additional modelling and assumptions (e.g. protein structure determined by x-ray crystallography). In other application, such as the calculation of solvation energies (ΔG^{solv}) by TI, multiple ensemble averages are combined in the calculation of a particular system property which may itself be compared to experimentally determined quantities. In both cases knowing the statistical uncertainty in ensemble averages due to finite sampling is critical in being able to determine the significance of the results.

It is important to preface any attempt to assess the convergence of an ensemble average calculated from an MD simulation with the irreducible aspect of the problem: one cannot know whether all the important regions of phase space have been sampled or not.[58-59] It is however possible to assess in a relative sense the quality of sampling for those regions of configurational space that have been sampled. A number of methods have been developed to do this based on the atomic positions sampled during a simulation.[60-64] One class of approaches involves a population analysis of bins corresponding to difference conformations in order to qualitatively assess the convergence of conformational sampling[63] or to estimate the effective number of independent samples.[62] Alternatively, principle component analysis based on position fluctuations have been used.[61, 64] However, all such methods come at a significant cost. In many cases they require storage of trajectories of configurations and subsequent reprocessing of the atomic coordinates. Another disadvantage of applications in which the atomic positions themselves are not of interest, is that these methods treat all positional degrees of freedom equally, irrespective of their influence on the property for which the ensemble average is being calculated. For example, during a calculation of the free energy of a ligand binding to a protein, the fluctuations of a protein loop which is remote to the binding site could dominate the assessment of convergence rather than the ligand:protein interaction itself. An alternative and much cheaper approach is to simply perform a statistical

analysis of the time series of values pertaining to the property of interest. This is the approach applied in this work.

The standard error of a sample mean i.e. the uncertainty in the estimate of the population mean from the mean of a sample, is given by:

$$SE = \frac{s}{\sqrt{n}} \tag{11}$$

where s is the standard deviation of the sample and n is the number of independent samples. The difficulty in calculating the standard error of a time series of a particular quantity obtained by MD simulation is that MD, by definition, produces correlated samples. In fact, even within a given system different properties may have correlations on different timescales. For example, the rotational autocorrelation time of a water molecule is on the order of 1-2 ps, local fluctuations within a lipid membrane can occur on the timescales of 10-100 ns, while large scale lipid-membrane motions require μs to sample adequately.

Autocorrelation analysis can be used to gain some insight into the number of independent samples, however this tends to be dominated by processes which occur on the shortest-timescales. The autocorrelation function is therefore not particularly useful in practical applications as the true number of independent samples obtained from an MD simulation is determined by the longest timescale fluctuations.[65-66] If the correlation time (τ) of the longest-timescale process is known, the number of independent samples can be estimated by dividing the total simulation time (t) by τ. Which in turn can be used to estimate the standard error:

$$SE \approx \frac{s}{\sqrt{\frac{t}{\tau}}} \tag{12}$$

where s is the standard deviation of the time series. However, the longest correlation time is rarely known *a priori* and there is no method known to this candidate that can be used to reliably calculate it for an arbitrary time series. The most common method used to estimate the standard error of a time series average is block averaging.[58] First published by Flyvbjerg and Petersen in 1989[65] the block averaging method implicitly deals with the fact that multiple correlations may be present on different timescales.[58] The method consists of dividing a trajectory of N snapshots into a series of blocks of sizes n, ranging from $n \approx 1$ to $n \approx N/10$. For each block size n, the standard deviation of the block averages is computed (s_n). The blocked standard error (BSE_n) given by:

$$BSE_n = s_n / \sqrt{N/n} \tag{13}$$

The standard error of the time series average is estimated from the value at which *BSE*(*n*) converges, shown graphically in Figure 2.1.

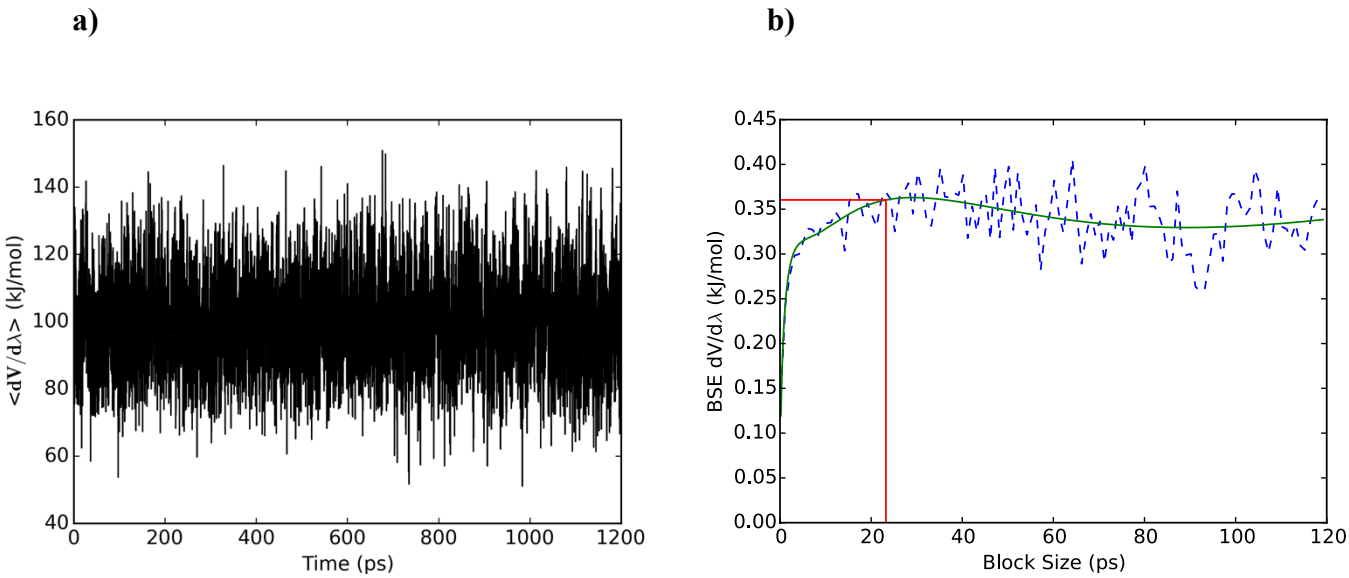

Figure 2.1 An illustration of the block averaging method: **a)** an example time series of $\partial V/\partial\lambda$, **b)** block averaging plot of the time series in panel **a)**. Green curve is a least square fit of Eq. 18 with m=2, red lines indicate where *BSE* has approximately converged.

The block averaging method has been implemented in many MD analysis packages[67-69], however there is no common method for computing the value at which *BSE*(n) has converged. For example, the Python MDAnalysis package[68] relies on human inspection of the curve to determine convergence, the GROMOS analysis toolkit[67] uses a numerical derivative based method to identify a turning point, while the GROMACS package[69] includes a method developed by Hess[70] that performs a least squares fit of the *BSE*(n) curve to a particular functional form based on a number of assumptions e.g. that the autocorrelation function is a sum of two exponentials. The lack of a consensus on how to algorithmically determine when *BSE*(n) has converged is indicative of the primary drawback of the method. Namely, that as the block size increases and the *BSE*(n) values become sensitive to the critical long-timescale fluctuations, the number of blocks decreases and thus the noise due to small sample size increases. This instability for large block sizes makes it difficult to robustly define whether the *BSE*(n) curve has converged and therefore the precise value of the block averaged standard error.

2.1.1 Kolmogorov–Smirnov Standard Error Estimate

In this work, we propose an alternative method to block averaging for estimating the standard error of a time series average. The method is based on the hypothesis that the standard error of a time

series average can be estimated from the distribution of fluctuations of the time series data itself. For this purpose we use the two-sample Kolmogorov–Smirnov (KS) statistic in which the equality between two sample distributions is determined from the maximum vertical distance between their normalised cumulative distributions.[71-72] The KS–statistic has a number of advantageous properties which make is suitable for this purpose: (1) it makes no assumptions about the nature of the distributions e.g. normality, (2) it involves no fitted parameters, (3) it is most sensitive to differences in the means and less sensitive to outliers. The standard error of a time series average can be estimated by multiplying the KS–statistic, obtained by comparing the 1st and 2nd halves of the time series, with the sample standard deviation (s) of the entire time series:

$$KS_{SE} = s \times \text{KS–statistic} \tag{14}$$

Note that since the standard deviation has the units of the time series values and the KS–statistic is a normalised quantity between 0 and 1, KS_{SE} (Eq. 14) has the same units as the time series property itself. The use of the KS–statistic is illustrated in Figure 2.2: **a)** is an example time series, **b)** shows the fluctuation distributions of the 1st (red) and 2nd (blue) halves of the time series data in **a)**. It is clear from both the frequency and cumulative frequency distributions in Figure 2.2 **b)** that there is a significant difference in the fluctuation distributions indicating a lack of agreement between the two portions of the time series.

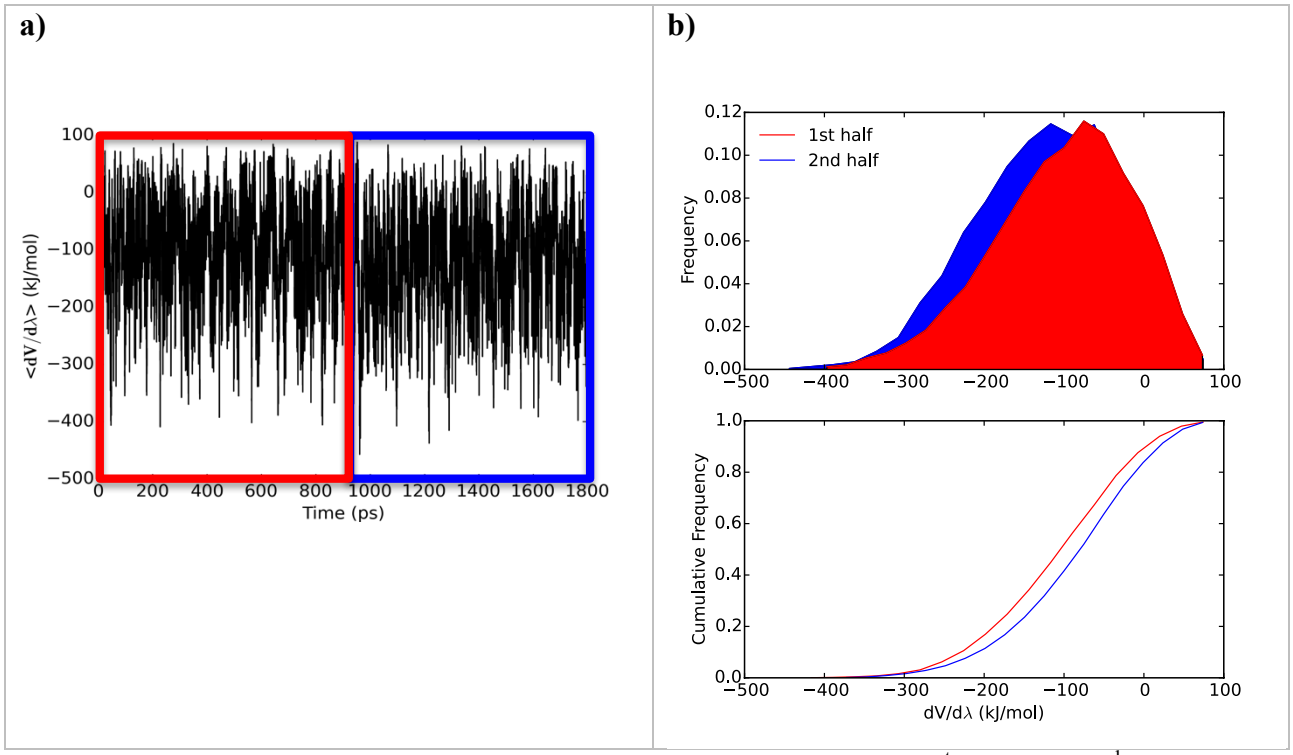

Figure 2.2 An illustration of the Kolmogorov–Smirnov statistic: **a)** the 1st (red) and 2nd (blue) halves of a $\partial V/\partial\lambda$ time series; **b)** the fluctuation distributions (top) and cumulative fluctuation distributions (bottom) for the 1st and 2nd halves of the time series shown in **a)**.

The following two sections describe steps to validate the KS_{SE} method for estimating the standard error of a time series average using an autoregressive model as well as extended time series data used in the calculation of ΔG^{solv} by TI.

2.1.2 Validation of KS Standard Error Estimate Using an Autoregressive Model

Simplified models are a useful way of demonstrating, validating and comparing computational methods as the relevant parameters can be directly controlled and the correct answer known *a priori*. In this case we generate time series data using a first order autoregressive model[73]:

$$X_t = c + \gamma X_{t-1} + \varepsilon_t(\sigma_\varepsilon) \tag{15}$$

where X_t and X_{t-1} are the current and previous values of the time series, c is a constant offset, γ is the "memory" parameter that controls the degree of coupling to the previous step, ε_t is white noise with standard deviation σ_ε. The utility of time series data generated from an autoregressive model is that the mean (μ), standard deviation (σ) and autocorrelation time (τ) can all be analytically expressed in terms of the model parameters:

$$\begin{aligned} \mu(X_t) &= \frac{c}{1-\gamma} \\ \sigma(X_t) &= \sigma_\varepsilon / \sqrt{1-\gamma^2} \\ \tau(X_t) &= -\frac{1}{log(\gamma)} \end{aligned} \tag{16}$$

An autoregressive model can thus be trivially parameterised to produce a time series with known statistical properties. Example time series generated with this autoregressive model are shown in Figure 2.3. The difference in the behaviour of the time series due to an increase in noise (σ) is illustrated by comparing the panels between the first and second columns, while an increase in the correlation time (τ) is illustrated between the rows. Note, the average value of these time series is clearly less well-defined for larger values of σ and τ. This is characteristic of differences in ensemble average properties calculated by MD simulation or the same property calculated on different systems.

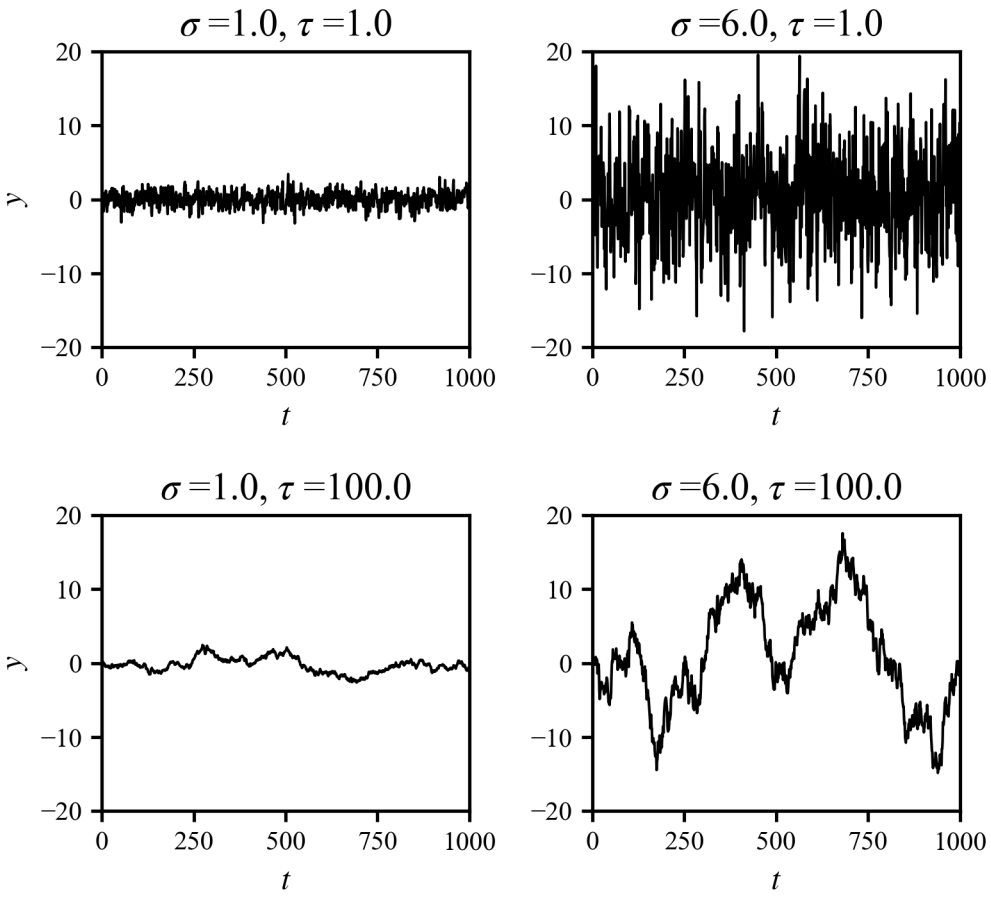

Figure 2.3 Example time series generated using an autoregressive model (Eq. 15) parameterised with different values of σ and τ; μ=0 in all cases.

Given that the underlying target mean of any time series generated with this model is predetermined, we can use the model to compare an estimated error in the average with the actual difference between the average calculated at a given value of t and the predetermined target mean. Note that even though the target mean is the maximum likelihood value of the calculated time series average, only in the limit of $t \to \infty$ will the two agree exactly. Since the value of τ has also been predetermined, we can compare a predicted error in the average to the standard error as determined by the number of independent samples according to Eq. 12. The KS–statistic standard error estimate (KS_{SE}) as a function of t can also be fitted using a least-squares method to a standard error model of the form:

$$f(x) = \frac{a}{\sqrt{x}} \tag{17}$$

Where a is the fitted parameter. Figure 2.4 shows a comparison of KS_{SE} (blue), the fit of a standard error model (Eq. 17) to the KS_{SE} curve (green dashes), the standard error as per Eq. 12 (black dots),

and the actual deviation from the underlying mean (black), for two different combinations of τ and σ. The underlying mean (μ) was set to 100 in both cases. From Figure 2.4 it can be seen that the KS_{SE} values and the fit of Eq. 17 to the KS_{SE} values (SE_{fit}) agrees well with the predefined properties of the time series. That is, the expected behaviour with respect to time (t), noise (σ) and correlation (τ) are all well reproduced by the KS_{SE} method. However, given the statistical nature of the problem, the utility of comparing individual time series is limited. A more relevant analysis is whether the KS_{SE} predicted error can accurately reproduce the standard error of a time series for a range of τ and σ values, at a given value of t. That is, how does the distribution of KS_{SE} predictions calculated from a large number of time series, compare to the predetermined properties of the time series. Also of interest is how the KS_{SE} predictions compare to the widely used block averaging method. However, in order to be able to perform such a large-scale distribution analysis on the block averaging method, an automated and robust method for algorithmically determining the block average convergence is required.

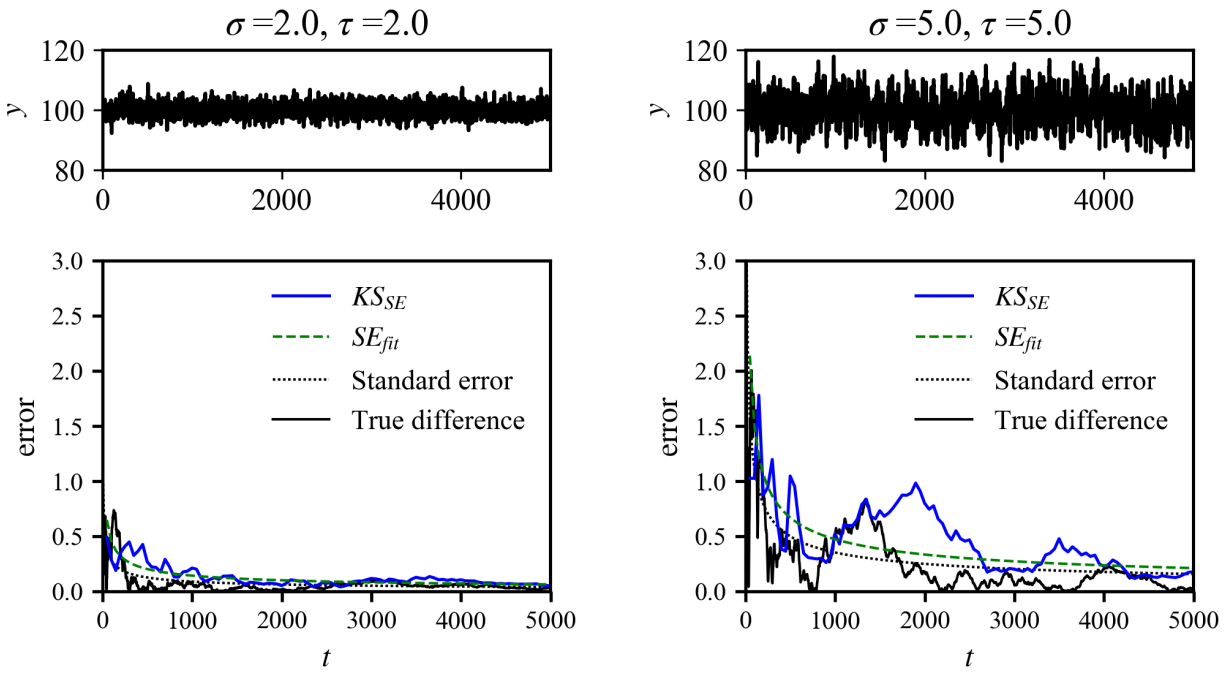

Figure 2.4 Example comparisons of KS_{SE} estimate (blue), a fit to the KS_{SE} values (green dashes), standard error (black dots) and the actual difference between target mean and the calculated average (black solid).

2.1.3 Automated and Robust Block Average Analysis

As noted previously, a range of algorithms to determine when a block average has converged have been proposed. The method implemented in the GROMOS analysis package based on numerical derivatives is unstable and does not return an error estimate in many cases. While the method of Hess[70] returns an error estimate in all cases, visual inspection of the fits it produces suggests the assumptions used to derive the fitting functional form frequently do not hold. The approach we have taken is similar to that of Hess[70], namely a least square fit of the *BSE*(*n*) curve. However instead of assuming a particular form of the autocorrelation function, we fit a general sum of exponential decay functions weighted by the square-root of the number of blocks for a given value of *n* i.e.:

$$BSE(n) = \sqrt{\frac{N}{n}}\left(\sum_{i=1}^{m} a_i e^{-b_i n} + c\right) \tag{18}$$

where *n* is the number of time series values per block, *N* is the total number of time series values, *m* is the number of exponential decay functions and a_i, b_i and *c* are the fitted parameters. Convergence was determined from the first root of the analytical derivative of Eq. 18. If the derivative of the fitted function had no roots (i.e. no turning points) the minimum value of the derivative was used to produce an error estimate. The result of this fitting approach is shown in Figure 2.1 **b)**, the fit to Eq. 18 for $m = 2$ is in green with the turning point indicated by the red lines.

While fitting to Eq. 18 offers significant improvements to the alternative methods tested, it still suffers from fitting issues. In many cases using a single exponential decay function led to a poor fit to the *BSE* curve and therefore the determination of convergence was unreliable. While in some cases the fit was significantly improved by the addition of a second exponential decay term, in others the use of a second exponential terms results in overfitting artefacts. However, the fits to data produced by a first order autoregressive model are well behaved for a single exponential term (i.e. $m = 1$ in Eq. 18). This scheme for determining convergence of the *BSE* curve could therefore be reliably used to perform large-scale distribution analysis on the block averaging method.

2.1.4 Distribution Analysis of KS_{SE} and Block Averaging methods

To perform a distribution analysis of the KS_{SE} and block averaging methods for estimates of the standard error of a time series average, we used an autoregressive model to generate *N* time series with a particular set of values for μ, τ, σ and t_{max}. Predictions made with these methods were then compared to the predetermined properties of the time series. That is, we compared the results to the actual distribution of deviations that were expected for a particular τ and σ at time t_{max}. These

distributions of difference between the target mean and the calculated average for a particular combination of τ and σ at time t_{max} represent the irreducible uncertainty of the problem. Since we have predetermined the correlation time (τ), we can also use Eq. 12 to estimate the true standard error. Such an analysis is shown in Figure 2.5, which contains the distribution of standard error predictions obtained using block averaging (orange), the KS_{SE} (blue) method, standard error according to Eq. 12 (black vertical line) and the true distribution of deviations (green). The distributions were generated with 4 different combinations of τ and σ from 500 time series at t_{max} =5000, and are represented with Gaussian kernel-density estimates. The results in Figure 2.5 show that the block averaging and KS_{SE} methods produce similar distributions which systematically over-estimate the standard error according to Eq. 12 (black vertical line). Note that the standard error line is identical to the median of the true distribution of deviations, within the uncertainty. Figure 2.5 also indicates that the two methods have a very similar dependence on σ, evident from comparing the top row ($\sigma = 5$) and bottom row ($\sigma = 10$) of panels in Figure 2.5. However, there is a slight difference in their dependence on τ. In the left column of panels ($\tau = 2$) the KS_{SE} method is further from the standard error line than block averaging, while in the right column of panels ($\tau = 5$) the KS_{SE} method is closer to the standard error.

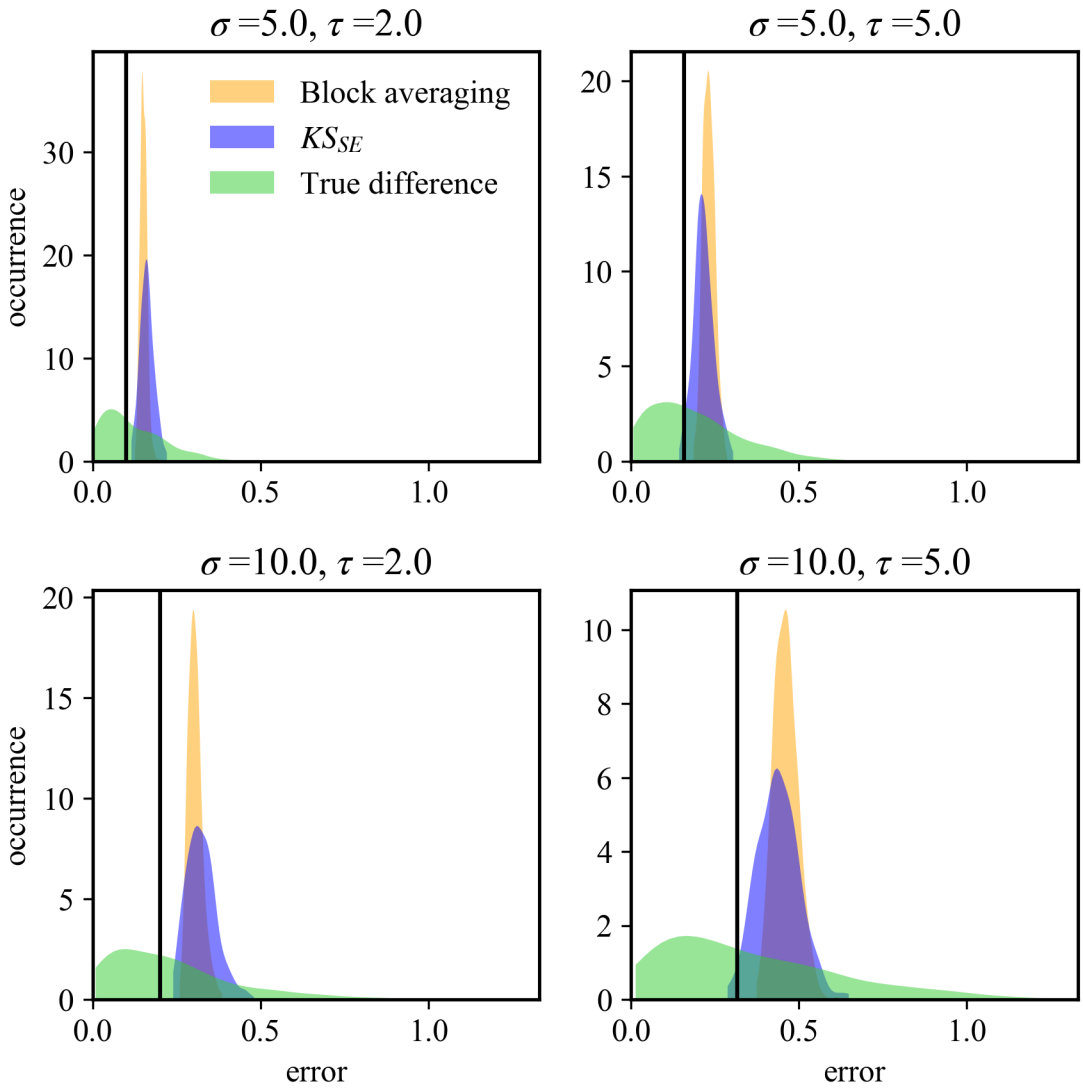

Figure 2.5 Comparison of distributions of error predictions for KS–statistic standard error (KS_{SE}, blue), block averaging (orange) and the true differences between target and calculated means (green). The classical standard error as predicted by Eq. 12 is shown as a vertical line. Distributions were generated from 500 time series with t_{max} =5000 and represented with Gaussian kernel-density estimates.

To further probe the dependence of the standard error predictions on the noise and correlation time, the distribution analysis was extended by calculating the mean and standard deviations of the predicted distributions for a wide range of τ and σ values. Figure 2.6 shows the dependence of the predicted standard error for σ between 1 and 20, for three values of τ (2, 10 and 20). The KS_{SE} and block averaging method results are shown in the left and right panels respectively. The distribution means are shown as solid lines with 1 standard deviation represented by the shaded areas below and above the mean. The standard error as per Eq. 12 for each combination of τ and σ at t = 5000 is shown as dashed lines. Figure 2.6 illustrates that for the range of τ and σ values shown, both

methods correctly account of the correlation and noise within time series data to produce reasonable estimates of the standard error. It also confirms the trend with respect to τ evident in Figure 2.5, namely that the KS_{SE} method produces systematically higher standard error estimates when the correlation is low (e.g $\tau = 2$) while block averaging produces systematically higher standard error estimates in high correlation cases (e.g. $\tau = 20$).

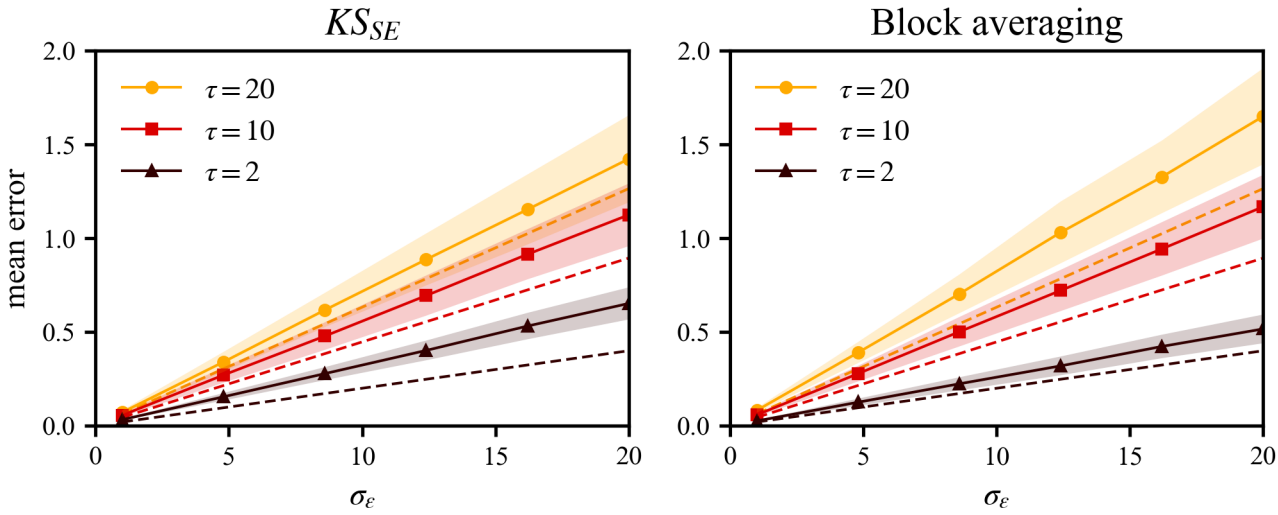

Figure 2.6 Extended distribution analysis. The dots interpolated with a solid line represent the mean of the distribution generated with particular τ and σ values for N=500 and t_{max}= 5000. Shaded areas represent 1 standard deviation from the mean.

The results above show that for a time series data generated using an autoregressive model the KS_{SE} method is an effective means to estimate the standard error of the time series average. The KS_{SE} method produced results which are very similar to block averaging without the added uncertainty of determining the convergence of *BSE* values.

2.1.5 Real Data Validation of KS_{SE} Predictions

While the simplified model discussed above offers insight into the behaviour and relative performance of the KS_{SE} method, it is also important to validate the KS_{SE} method against actual MD time series data. For this purpose, 10 ns trajectories of $\partial V/\partial\lambda$ (used in the calculation of ΔG^{solv}) were analysed from simulations of 26 all atom hydrocarbons solvated in SPC water. The values of $\partial V/\partial\lambda$ were stored every 10 ps thus producing a time series of 100,000 data points. The simulation setup was the same as described in section 1.3.4. The first 1 ns of data was omitted to remove any bias from the initial conditions. The last 8 ns of the trajectory (90,000 data points) was used to

calculate a reference time series average, which was compared to the average and standard error estimate calculated from the 1–2 ns portion of the data (10,000 data points). That is, the reference error was taken to be the difference between the average calculated from the 1–2 ns and 2–8 ns portions of the data. The results for block averaging and KS_{SE} methods are summarised in Table 2.1. As was observed in the autoregressive model validation results, the performance of the block averaging and KS_{SE} methods are very similar. With the differences between the two methods in Table 2.1 being well within the statistical uncertainty.

Table 2.1 Summary of time series average standard error predictions for the block averaging and KS_{SE} methods applied to $\partial V/\partial \lambda$ time series data. All values, excluding the last column, are in kJ mol^{-1}.

	range	AE[a]	AUE[b]	max. error	*N* under est.
Block Average	0.05–0.43	-0.07 ± 0.18	0.13 ± 0.14	0.63	13
KS_{SE}	0.07–0.32	-0.07 ± 0.19	0.14 ± 0.15	0.63	14

[a]average error
[b]average unsigned error

In this and the previous section we have shown that the KS_{SE} method performs similarly to the block averaging method for estimating the standard error of a time series average. The primary advantage of the KS_{SE} method is its robustness. The difficulty in applying the block averaging method, particularly in an automated protocol, stems from the fact that that as the *BSE* values converge to the standard error, the number of blocks (*N/n* in Eq. 18) decreases. This decrease in the number of blocks causes an increase in the statistical fluctuations in *BSE* values (due to taking the standard deviation of small number of values) thus making it difficult to fit in a general and robust manner. The KS_{SE} method does not suffer from the same drawbacks, and as we will demonstrate in the following section, a simple extension of the KS_{SE} method allows for the calculation of the equilibration time (due to unrepresentative initial conditions). This method is also used to derive a simple heuristic to assess the robustness of the standard error estimate.

2.1.6 Estimation of Equilibration Time

While formally all configurations of a system are valid states of the equilibrium ensemble, sampling limitations dictate that for equilibrium properties to be accurately reproduced by simulation, they must be dominated by low-energy states which can be adequately sampled on the time scale of the simulation. These low-energy states must appear in a simulated equilibrium ensemble with the

correct probabilities in order for the calculated properties to accurately reproduce true physical properties of the system. Therefore, configurations of a sampled equilibrium ensemble which are not representative of the low-energy states—for example those heavily biased by the initial conditions—can reasonably be omitted from the calculation of equilibrium properties. The time that a system takes to relax from an unfavourable set of initial conditions is often referred to as the equilibration time (t_{eq}).

The KS–statistic estimate of the standard error (KS_{SE}, Eq. 14) described in section 2.1.1, can also be used to estimate t_{eq} caused by unrepresentative initial conditions. Given that all unrepresentative configurations originate from the system state at $t = 0$, t_{eq} is estimated by progressively excluding data from the beginning of the time series and plotting the resulting KS_{SE} values as calculated by Eq. 14. Specifically, from the two-sample KS–statistic obtained between the 1st and 2nd halves of the interval $t = [t_{excl}, t_{max}]$ where t_{excl} is sequentially increased from 0 to t_{max}:

$$\begin{aligned} F(t_{excl}) &= KS_{SE}([t_{excl}, t_{max}]) \\ t_{excl} &\in [0 \rightarrow t_{max}] \end{aligned} \tag{19}$$

Figure 2.7 shows the results of this scheme (lower panels) for two example time series (upper panels). The example in the left panels of Figure 2.7 show the analysis for a $\partial V/\partial \lambda$ time series in which the initial conditions were representative of the system at equilibrium ($t_{eq} = 0$), in this case the KS_{SE} values corresponding to the entire time series being included (i.e. t_{excl} close to 0) are small and progressively increase as more of the time series is omitted. This increase in KS_{SE} values is due to the decrease in sample size and the corresponding increase in uncertainty. The example in the right panels of Figure 2.7 is of a case where the initial conditions do not represent the system at equilibrium ($t_{eq} >> 0$). This example is of a pressure (P) time series generated from an NPT liquid phase MD simulation that was initialised with inappropriate box dimensions. The upper right panel in Figure 2.7 shows that the system takes ~400 ps for the pressure to equilibrate. This equilibration period is clearly represented in the lower right panel in Figure 2.7, as an initial drop in the KS_{SE} values for t_{excl} close to 0. That is, in contrast to the $t_{eq} = 0$ case in which the KS_{SE} values initially increased as $t_{excl} \rightarrow t_{max}$, the KS_{SE} values initially decreases to a minimum value at $t_{excl} \approx 400$ ps, after which it increases as expected for smaller sample sizes. The value of t_{eq} is shown with a red vertical line in Figure 2.7.

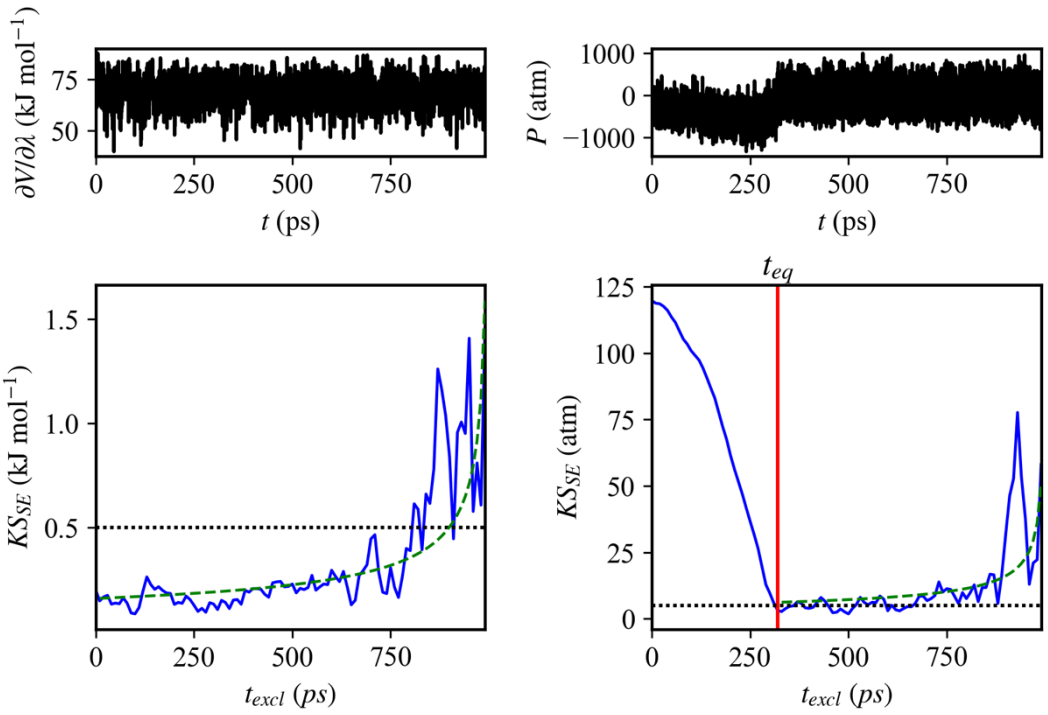

Figure 2.7 Two examples of the KS_{SE} method for detecting equilibration region within a time series. Upper panels are of the time series data points. Lower panels show the behaviour of KS_{SE} as the amount of excluded data (t_{excl}) is increased.

The detection of the equilibration time is done algorithmically by predefining a target error (E_{target}) for the time series average (shown as black dots in Figure 2.7) and comparing the KS_{SE} values relative to the target error. Specifically, if the KS_{SE} value is greater than the target error for $t_{excl} = 0$, and the KS_{SE} value is less than the target error for some value of $t_{excl} > 0$, t_{eq} is taken to be the first value of t_{excl} such that $KS_{SE} < E_{target}$ i.e.:

$$t_{eq} = \begin{cases} \min(t_{excl}) \text{ where } KS_{SE} < E_{target} & \text{, if } KS_{SE}^{t_{excl}=0} > E_{target} \\ 0 & \text{, otherwise} \end{cases} \tag{20}$$

Note that since the KS_{SE} values have the same units as the time series property being averages, E_{target} represents the error tolerance relevant to a given application. Data points within the equilibration region of a time series ($t < t_{eq}$) can then be excluded from all ensemble average calculations to avoid bias from the initial conditions.

Note that a more precise determination of t_{eq} can be obtained from the minimum KS_{SE} value rather than the minimum t_{excl} where $KS_{SE} < E_{target}$. However, given the statistical noise in the KS_{SE} values, such a method would require smoothing or fitting of the data. Another potential improvement for determining t_{eq} more precisely is by applying the KS_{SE} analysis (Eq. 19) as a sliding window rather

than over the entire time series. In the approach described above, the value of t_{eq} is to some degree dependent on the length of the time series as the contribution of unrepresentative values to the overall distribution is smaller for longer time series. This effect would be removed by using a fixed-length sliding window as each value of the time series would contribute equally to the overall distribution. However, it is not clear how best to determine an appropriate window size for a given time series. While these extensions may warrant the added complexity in some applications, to date we have not found them necessary.

2.1.7 Convergence Robustness Heuristic

The KS_{SE} approach summarised in Eq. 19 can also be used to quantify the robustness of the standard error prediction with a simple heuristic. This involves using a predefined target error (E_{target}) for the time series average, and comparing the values of KS_{SE} to this target error as $t_{excl} \rightarrow t_{max}$. If the values of KS_{SE} are below E_{target} for the majority of the time series, this strongly suggests that the target error has been met. Conversely, if KS_{SE} values never drops below E_{target} or does so only briefly, this suggests that the E_{target} has not been met. This is quantified by a convergence robustness score (CR_{score}) which is defined as the ratio of the portion of the time series which is continuously below E_{target} and that portion of the time series which is above (or sporadically crosses) E_{target}. The division of the time series with respect to $KS_{SE} < E_{target}$ is performed on the fit of the KS_{SE} values to a standard error model (Eq. 17). This scheme for quantifying the robustness of convergence with respect to a target error is illustrated in Figure 2.8. The left panels show an example time series (top) and corresponding convergence analysis (bottom) for a well converged case, the target error was 1 kJ mol^{-1} and CR_{score} was 15.7. While the example shown in the right panels of Figure 2.8 is not as well converged to the same target error and has a CR_{score} of only 0.4.

This convergence robustness heuristic (CR_{score}) offers a rigorous control on the confidence of the error estimate made using the KS_{SE} method. It allows for explicit control over the trade-off between efficiency and robustness and can also be trivially incorporated into automated calculation protocols.

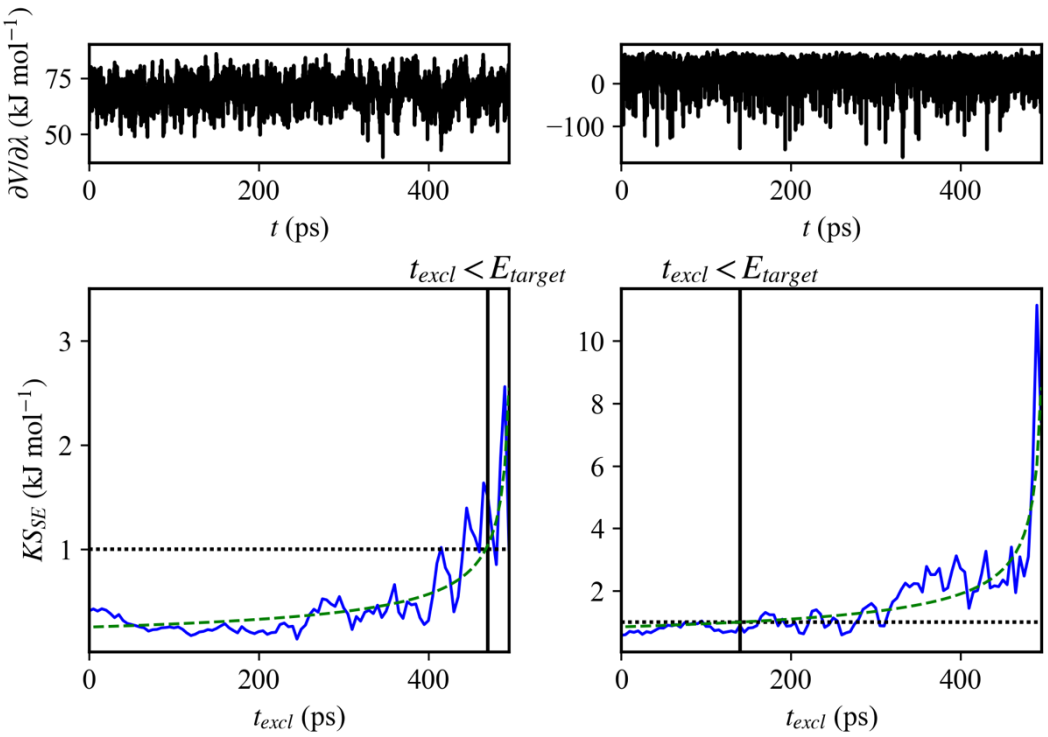

Figure 2.8 Illustration of convergence robustness analysis on two example time series. Upper panels are of the example $\partial V/\partial\lambda$ time series data used in the calculation of ΔG^{solv} by thermodynamic integration. Lower panels are of KS_{SE} values (blue) and corresponding fit to a standard error model (Eq. 17, green dashes) as calculated by Eq. 19. The target error of 1 kJ mol^{-1} is shown in black dots.

The KS–statistic based methods described above is applied extensively as a part of an automated protocol for the calculation of ΔG^{solv} by TI developed as part of this thesis. In combination with the numerical integration error analysis (presented in the next chapter) the KS_{SE} method has been used to quantify ensemble average uncertainty in the calculation of over 1200 ΔG^{solv} values which consisted of over 14500 ensemble averages calculations. A description of the automated ΔG^{solv} calculation protocol along with an analysis of its performance is provided in section 2.3. While the application of the method to date has been fairly limited in terms of the range of properties, the method is appropriate for estimating the standard error in any property which can be represented as a time series average.