LU TP 22-44
June 2022

Goodness-of-fit Tests for Time Dependent Ensemble Averages

Jake Lumsden

Department of Astronomy and Theoretical Physics, Lund University

Supervised by Dr. Tobias Ambjörnsson

LUNDS UNIVERSITET

Abstract

Fitting a model to a time-dependent ensemble average is a process repeated frequently throughout biophysics. A selected ensemble-averaged observable ($\langle y(t) \rangle$) for a given system can be predicted through the use of an estimated ensemble average, where the estimated ensemble average is created via simulated or experimental data sets. Fitting a model to this estimated ensemble average allows for estimations of $\langle y(t) \rangle$.

Often, one tests the quality of a fitted model through the use of a *'goodness-of-fit'* (GOF) procedure. The quality of the model is determined by the placement of a test statistic (S) on its associated probability distribution ($\phi(S)$). Traditional choices of S, such as the normalised residual sum of squares (RSS), neglect correlations in fluctuations of the ensemble average around the fitted model. Under this assumption, the normalised RSS is distributed according to the χ^2-distribution ($\phi_{\chi^2}(S)$), with mean ($\boldsymbol{\mu}$) and variance (σ^2) proportional to the degree of freedom (υ) of the fitted model. The inability of the traditional χ^2-GOF procedure to account for these correlations can lead to less reliable evaluations of the quality of a fitted model.

The thesis covers the derivation and validation of the correct form of $\phi(S)$ when correlations are considered, for use in a new GOF procedure. The new GOF procedure was tested under varying parameters, correlation types and ensemble make-ups. Testing environments included three ensemble generating prototype models, and three movies of noisified simulations of vesicle movement. It is demonstrated that the new GOF procedure correctly accepts and rejects well and poor fitting models respectively, and is a valid indicator of model quality. Furthermore, it is shown that compared to the traditional χ^2-GOF procedure, the new GOF procedure is a more accurate measure of model quality under a variety of correlation types, is reliable in a greater region of parameter space, and performs better in all tested scenarios.

Popular Science - No Correlation Left Behind

Studies of micro and nanoscopic particles and molecules have been of huge benefit to the biophysics community. The ability to conduct experiments at the microscopic level has generated new knowledge about once unseen biological processes from virus incubation to the life cycle of bacteria.

At the microscopic level, the recorded trajectories of a given system of particles are often grouped into what is referred to as a time dependent ensemble average. A time dependent ensemble average describes a given system of particles as a single stream of time dependent data, that data being a chosen metric by which to average over time. Time dependent ensemble averages allow for easier and more reliable interpretation of a system of particles, and for the extraction of certain system dependent parameters, such as how and the rate at which certain particles move under a set of pre-defined conditions.

Fitting a model to the time dependent ensemble average allows for predictions further in time to be made, and further parameters to be extracted, such as the rate of diffusion.

In practice, one often estimates a chosen ensemble-averaged observable ($\langle y(t) \rangle$) through the use of an estimated ensemble average made up of simulated or experimental data sets containing $\langle y(t) \rangle$. Fitting a model to this estimated ensemble average then allows for the extraction of a prediction of $\langle y(t) \rangle$.

For both predictions and extracted parameters to be accurate, one must make sure that the model is well fitting. Goodness-of-fit (GOF) procedures are used frequently throughout various scientific communities to ensure that models are well fitting to their ensemble counterparts. Traditional GOF procedures, such as the χ^2-GOF procedure, neglect any correlation among the fluctuations of the ensemble average around the fitted model, leading to less reliable determination of a fitted model's quality.

This thesis fills this gap in traditional GOF procedures, developing a new GOF procedure which includes the correlations in fluctuations of a ensemble average around a fitted model. The new GOF procedure will allow scientists within the biophysics community to make more reliable predictions and extract more accurate parameters from a given time dependent ensemble average. The hope is that researchers will take to and use this new GOF procedure to further the knowledge of intricate biological particles and processes, for example, virus structure and transmission, or the diffusion of molecules through the lipid membrane.

Acknowledgements

I'd like to thank Tobias Ambjörnsson, whose continuous ideas and guidance enabled me to overcome any problem encountered throughout the thesis. I'd also like to thank Marc Pielies Avellí, whose feedback on various drafts was of great help.

Contents

List of Figures

List of Tables

1 Introduction

Time dependent ensemble averages are employed frequently throughout multiple fields of physics, particularly within biophysics. Studies in which ensemble averages are common place include virus tracking [1, 2, 3], membrane dynamics [4, 5, 6], fluorescent protein tracking [7, 8, 9], and more generally, single particle tracking as a whole [10, 11, 12].

An ensemble-averaged observable ($\langle y(t) \rangle$) is a theoretical value gained by averaging a given $y(t)$ over an infinite number of initially identical systems. For example, $\langle y(t) \rangle$, could be the ensemble-averaged mean squared displacement (MSD) or mean velocity of a particular system of particles. In practice, estimates of $\langle y(t) \rangle$ are obtained through averaging over a finite number (M) of systems, either generated through simulation or experimentation. From this point on, 'ensemble average', is used to describe a estimated ensemble average of a given system.

The next step is often to then fit a model to the ensemble average. The fitted model can then be used as estimator of a given $\langle y(t) \rangle$ further in time, or used to extract subsequent parameters related to $\langle y(t) \rangle$. Physicists in this position are often faced with the *'model selection problem'* [13, 14], that problem being, which model does one fit to the ensemble average that will yield the most reliable parameters. At this stage, one would often turn to an aptly named *'Goodness-of-fit'* (GOF) procedure in order to gauge the quality of a given model's fit to an ensemble average. In general, the GOF procedure involves the calculation of a test statistic (S), and the evaluation of its associated distribution ($\phi(S)$) [15, 16, 17]. The calculated S for a given system can then be compared to its position on the corresponding $\phi(S)$, and used to determine the quality of the fitted model. If S lands in the upper (tail) end of $\phi(S)$, the model is deemed poor.

The method one uses to fit a function to a given ensemble average can be separated into two main schools, regression (linear and non-linear), and Bayesian methods (Appendix E). Popular regression approaches, such as weighted-least-squares (WLS), fit a model to a given ensemble average by minimising a chosen cost function [13, 18]. Bayesian methods offer an alternative, selecting a model by maximising the appropriate likelihood function [14, 16, 19].

Oftentimes, GOF procedures for testing the quality of a fitted model assume the fluctuations of an ensemble average around a fitted model ($\mathbf{\Lambda}$) are independent and identically distributed (i.i.d). Here, $\mathbf{\Lambda}$ is a vector with length equal to the number of sampling points (N), with each entry equal to the residual at a given N, where a residual is defined as the difference between the ensemble average and fitted model. It can be demonstrated that the correlations among the components of $\mathbf{\Lambda}$ are in general not i.i.d (Section 5.1). If a given position along the ensemble average is above the fitted model, then in practice it is likely the position at the following time step will continue this trend [13]. The lack of consideration of correlations among the components $\mathbf{\Lambda}$ in current GOF procedures can lead to poor estimates of model fit quality.

In cases where the number of trajectories (M) is large, the multivariate central limit theorem (CLT) proposes that the distribution of $\mathbf{\Lambda}$ ($\varrho(\mathbf{\Lambda})$) can be represented by a multivariate distribution. Therefore, around a well fitting model, the distribution of residuals $\varrho(\mathbf{\Lambda})$ should conform to multivariate normality [13]. Many suitable approaches to test for multivariate normality exist [15], in this thesis the approach is based on the χ^2-GOF procedure.

In the χ^2-GOF procedure, an S calculated from the residuals is used to give a measure of the GOF of a given model. In the case that the components of $\mathbf{\Lambda}$ are uncorrelated, one would expect an S of approximately N [20], with an S greater than N denoting a poor fitting model.

Formally, one uses an S referred to as the residual sum of squares (RSS). In this case, S is a sum of $\mathbf{\Lambda}^2/s$, where $s = \sigma/\sqrt{M}$, with σ denoting the standard deviation of the components of $\mathbf{\Lambda}$. If the components of $\mathbf{\Lambda}$ are indeed i.i.d., one would expect $\mathbf{\Lambda}^2/s$ to be of the order one, and to be described by the χ^2-distribution ($\phi_{\chi^2}(S)$) with mean, $\boldsymbol{\mu}$, and variance, σ^2, where $\boldsymbol{\mu} = N$ and $\sigma^2 = 2N$ respectively [20].

The GOF of a given model using this approach can be evaluated using $\phi_{\chi^2}(S)$, where $\phi_{\chi^2}(S)$ has a $\boldsymbol{\mu}$ and variance σ^2 determined by the degree of freedom (υ), where $\upsilon = N$. This process is referred to as the χ^2-GOF procedure [20]. Due to this, the χ^2-GOF procedure cannot account for any correlations among the components of $\mathbf{\Lambda}$, which would alter σ^2. This inability of the χ^2-GOF procedure to account for correlations in $\mathbf{\Lambda}$ can lead to poor evaluations of model quality, which in turn leads to less accurate predictions of $\langle y(t) \rangle$.

In the correlated cases, current approaches, such as the Hotelling's T-squared test statistic (T^2) (Section 3.4) and it's associated distribution, can prove problematic in their application. The dimensionality of a given system can cause inaccurate T^2 to be calculated, leading to poor, or in certain cases, totally unreliable estimates of the GOF of a given model (Section 3.4) (Figure I).

This thesis concerns the development of a new GOF procedure which considers the correlations among the components of $\mathbf{\Lambda}$. This new GOF procedure employs methods demonstrated by Gil-Pelaez, Imhof and Davies to evaluate a $\phi(S)$ which includes correlations in the components of $\mathbf{\Lambda}$ [21, 22, 23]. The proposed GOF procedure aims to bring more accurate and reliable GOF testing to scientific communities where fitting models to ensemble averages is common practice, leading to more reliable modeling and parameter estimation.

In order to evaluate $\phi(S)$, one must invert the characteristic function (CF) (Fourier transform of $\phi(S)$) through the use of the inverse Fourier transform. Gil-Pelaez provides the proof that the inversion theorem presented by Lévy is capable of inverting a given CF, producing an evaluated ϕ for use as part of a GOF procedure [21]. Imhof demonstrated the validity of Gil-Pelaez's proof, by successfully using numerical methods to evaluate the theorem and invert a given CF [22]. This is followed by work from Davies, who provides a demonstration that the inversion theorem can be suitably applied to a χ^2-CF and evaluated to provide a stable ϕ [23]. Modern approaches have turned to approximate methods

to evaluate a required ϕ. Approximate methods based on moment evaluation have been demonstrated by Solomon et al [24], on cumulants by Lui et al [25], and using Fast Fourier Transforms (FFTs) by Witkovsky [26]. Duchense et al provide a comprehensive review of the method presented by Lui et al, and demonstrate its shortcomings when compared to the exact methods demonstrated by Imhof and Davies [27]. Bodenham et al provide further comparisons between approximate and exact methods, concluding that in offline situations exact methods are more favourable for the evaluation of a given ϕ [28], with particular emphasis on the approaches demonstrated by Imhof and Davies.

The primary testing and validation methods for the new GOF procedure employ the use of trajectories generated from three prototype models (Appendix B). These prototype models include Brownian motion (BM), fractal Brownian motion (FBM) and continuous time random walks (CTRWs). These three prototype models are of particular importance within the biophysics community, where BM is commonly used to describe diffusive cellular processes [29,30], FBM crowding dynamics [31, 32], and both FBM and CTRWs have been of use in describing motion within cell membranes [33, 34]. The three prototype models, aside from generating test trajectories, allow for the creation of validation procedures, through repeated simulation and manual evaluation of a given prototype model's 'true' $\phi(S)$ ($\phi(S)_{true}$). These three prototype models serve as appropriate examples of biophysical processes, but the new GOF procedure can be applied to any given system, with a model fit using an arbitrary choice of fitting method. On that note, the predominant fitting method used throughout this thesis is the weighted-least-squares including correlation error (WLS-ICE) procedure, developed and validated by Fogelmark et al [13]. The WLS-ICE method is a regression based approach, fitting a model to a given ensemble average through minimisation of a χ^2 cost function.

In the following section, the scope of the thesis is highlighted. Section 3 defines the selected S and details the underlying mechanisms used in the new GOF procedure. Section 4 describes the inversion and integration methods as well as the make-up of the estimate ensembles used for testing throughout the thesis. Section 5 presents the results of the testing of the new GOF procedure. Section 6 provides instruction on the use of $\phi(S)$ to determine the GOF of a given model. Section 7 concludes the thesis. Additional tables, images and detail can be found in the relevant Appendices.

2 Problem Outline

This section describes the problem outline, highlighting the scope of the thesis. The steps of the new GOF procedure and their impact on the model selection problem are detailed. The sequence diagram is depicted in Figure 1. An example of $\langle y(t) \rangle$, with a fitted model ($f(\theta)$), and associated $\mathbf{\Lambda}$ is presented in Figure 2. Here, $f(\theta) = \theta t^{\theta}$, where θ is a fitting parameter calculated via the WLS-ICE method and $t = (1, ..., N)$.

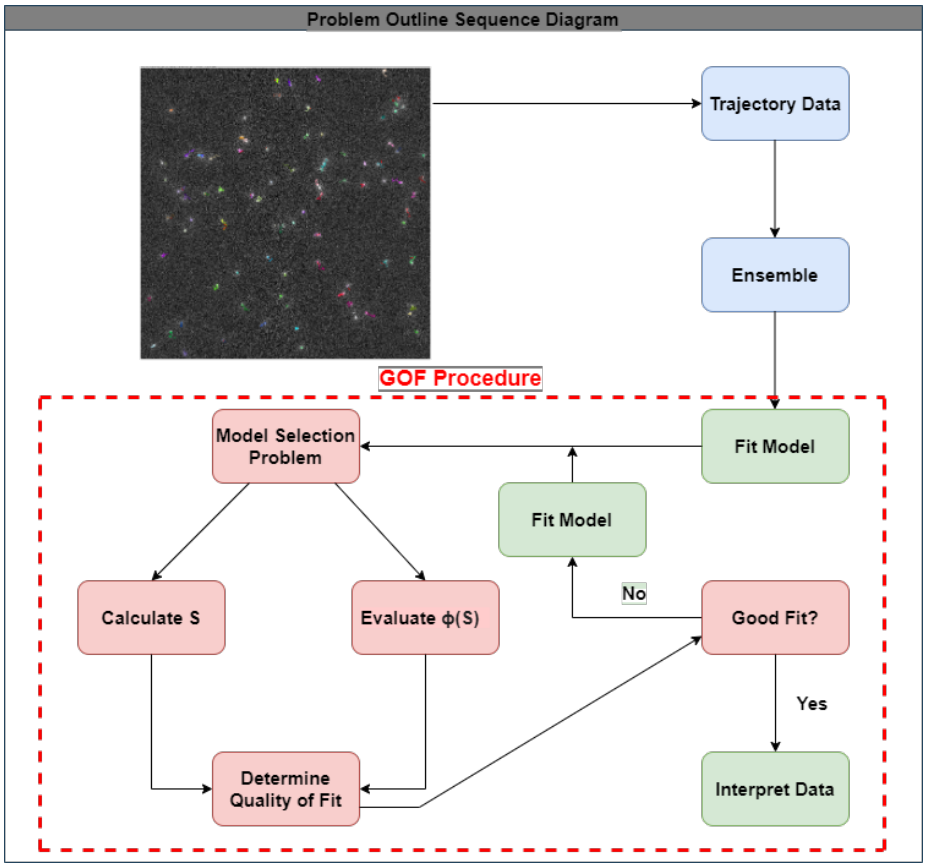
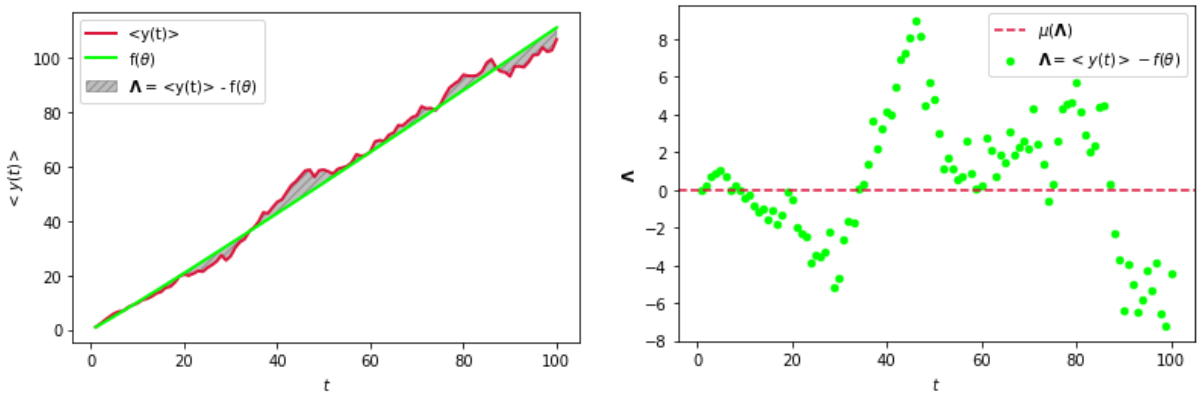


Figure 1: **Problem Outline - Sequence Diagram** Figure 1 provides a step-by-step sequence of the fitting of a model to an ensemble average and subsequently determining its GOF. The first stage (blue) concerns the collection of trajectories from experimental data, and the creation of a suitable ensemble average. The second stage (green) denotes steps that are associated with model fitting. The third section (red) lists the steps associated with the model selection problem. The area sectioned off with red dashes covers the steps included within the new GOF procedure.

In the steps prior to fitting (Figure 1, blue boxes), trajectories are extracted from experimental or simulated data, compiled and transformed into an ensemble average of a selected time dependent observable, $\langle y(t) \rangle$ (Figure 2) . Note, that throughout this thesis, all ensemble averages used are estimates of an analytically exact ensemble average. This is done by averaging over a finite number of M trajectories with N sampling points to create an estimated ensemble average of N sampling points. Though crucial, this stage has no effect on the outcome of the new GOF procedure, rather only the ensemble average to which the model is fit.



(a) **Fitted model. Shaded Residuals.** (b) **Extracted Residuals**

Figure 2: $\langle y(t) \rangle$ **vs.** N. Figure 2 provides an example of $\langle y(t) \rangle$ over N sampling points, with $f(\theta)$ and associated $\mathbf{\Lambda}$. Figure 2 (a) shows $f(\theta)$ fitted to $\langle y(t) \rangle$, with the shaded area being the difference between $\langle y(t) \rangle$ and $f(\theta)$. Figure 2 (b) provides a plot of the $\mathbf{\Lambda}$ extracted from the shaded area of Figure 2 (a). $\langle y(t) \rangle$ was generated using the BM prototype model, with $M = 100$ trajectories and $N = 100$ sampling points. Note, that in Figure 2 (b) $\boldsymbol{\mu}(\mathbf{\Lambda}) = 0$.

After the creation of a suitable estimate ensemble average, the following step (Figure 1, green boxes) is to fit a model (Figure 2). The outcome of this stage varies depending on the fitting method used and amount of data modelled. Though choice of fitting method will alter the model fit (Figure C), this thesis does not concern finding the optimal fitting method, rather focusing on providing a more reliable solution to the model selection problem.

On that note, the subsequent steps (Figure 1, red boxes) concern the model selection problem. The new GOF procedure aims to solve this problem via calculation of a suitable S and $\phi(S)$.

The region of Figure 1 outlined by red dashes provides an outlook of the GOF procedure as a whole, concerning both the fitting of a model, and the solving of the model selection problem.

To summarise, this thesis will aim to provide a new, more reliable solution to the model selection problem through creation of a new χ^2-based GOF procedure. In which a model using an arbitrary choice of fitting method can be fit to an ensemble average, the GOF procedure can then calculate S and evaluate its associated $\phi(S)$. S can then be used to determine the acceptance or rejection of a model, given its position on $\phi(S)$. At this point parameters can be extracted, or a new model fit, depending on acceptance or rejection of the fitted model respectively.

3 Theory

This section details the underlying mechanisms of the new GOF procedure. Firstly, the multivariate CLT and $\varrho(\mathbf{\Lambda})$ are defined. Next, covariance ($\mathbf{\Sigma}$) and Pearson correlation ($\mathbf{P}$) matrix estimation and the selection of a suitable S is discussed. Finally, the use of CFs as a method of evaluating $\phi(S)$, and the $\boldsymbol{\mu}$ and σ^2 of both $\phi(S)$ and $\phi_{\chi^2}(S)$ is detailed.

It is important to draw a distinction between analytically exact variables, meaning those derived from an ensemble average, where the average is take over an infinite number of systems, and those estimated via a finite number of simulated or experimental data sets. Moving forward, $(...)^*$ represents an analytically exact variable or observable.

3.1 The Multivariate Central Limit Theorem

The new GOF procedure relies on the multivariate CLT to approximate $\varrho(\mathbf{\Lambda})$. This section first defines the multivariate CLT, with the following (Section 3.2) detailing its application to approximating $\varrho(\mathbf{\Lambda})$.

Let us define a sample vector of N sampling points,

$$\mathbf{Y}^{(m)} = (y_1^{(m)}, ..., y_N^{(m)}), \tag{3.1}$$

where m denotes a given trajectory, and $m = (1, ..., M)$. We can then define the mean at a given n, with $n = (1, ..., N)$,

$$\bar{y_n} = \frac{1}{M} \sum_{m=0}^{M} y_n^{(m)}, \tag{3.2}$$

where $\bar{(...)}$ denotes the mean of a given observable.

Let us now define a new sample vector containing N samples of $\bar{y}$,

$$\mathbf{Y} = (\bar{y_1}, ..., \bar{y_N}), \tag{3.3}$$

where $\mathbf{Y}$ is from a random sample of a multivariate distribution. If we denote by γ the mean of $\mathbf{Y}$ and by $\mathbf{Q}^*$ the covariance matrix for the different components of $\mathbf{Y}$, then, according to the multivariate CLT,

$$\sqrt{M}(Y - \gamma) \longrightarrow N[0, \mathbf{Q}^*], \tag{3.4}$$

as $M \longrightarrow \infty$. The multivariate CLT thus proposes that $\phi(\mathbf{Y})$ can be approximated by a N-dimensional multivariate Gaussian [35, 36],

$$\phi(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{N}{2}} \mid \mathbf{\Sigma}^* \mid^{\frac{1}{2}}} e^{\left(-\frac{(\mathbf{Y}-\gamma)' \mathbf{\Sigma}^{*-1} (\mathbf{Y}-\gamma)}{2}\right)}, \tag{3.5}$$

where $(...)'$ denotes the transpose, $\mid ... \mid$ the determinant of a given observable, and $\mathbf{\Sigma}^*$ is [35],

$$\mathbf{\Sigma}^* = \frac{\mathbf{Q}^*}{M}. \tag{3.6}$$

Note that Equation 3.5 only holds when M is sufficiently large.

3.2 The Distribution of Fluctuations

Following on from Section 3.1, the multivariate CLT can used to represent $\varrho(\mathbf{\Lambda}, \theta)$. Rather than a random sample vector, treating $\mathbf{Y}$ as $\mathbf{\Lambda}$, where numerically,

$$\mathbf{\Lambda} = (\mathbf{y} - \mathbf{f}(\theta)), \tag{3.7}$$

where $\mathbf{y}$ are points along the ensemble average and $\mathbf{f}(\theta)$ the fitted model, with J free fitting parameters $\theta = (\theta_1,...,\theta_J)$. One can then rewrite Equation 3.5 to describe $\varrho(\mathbf{\Lambda}, \theta)$, giving an altered multivariate normal distribution,

$$\varrho(\mathbf{\Lambda}, \theta) = \frac{1}{(2\pi)^{\frac{N}{2}} \mid \mathbf{\Sigma}^* \mid^{\frac{1}{2}}} e^{\left(-\frac{(\mathbf{\Lambda}-\boldsymbol{\mu})' \mathbf{\Sigma}^{*-1} (\mathbf{\Lambda}-\boldsymbol{\mu})}{2}\right)}, \tag{3.8}$$

where $\boldsymbol{\mu}$ is the mean difference,

$$\boldsymbol{\mu} = \langle \mathbf{y} - \mathbf{f}(\theta) \rangle. \tag{3.9}$$

3.3 Covariance and Correlation Matrix Estimation

In certain systems, such as in the BM prototype model (Appendix B), $\mathbf{\Sigma}^*$ can be derived analytically [13, 37], in the general case however, $\mathbf{\Sigma}$ must be estimated.

The unbiased estimator of $\mathbf{Q}$ for a collection of $\mathbf{Y}^{(m)}$ is given by multiplying by $\mathbf{Y}'^{(m)}$ and averaging over the sample [38],

$$\mathbf{Q} = \frac{1}{M-1} \sum_{m=1}^{M} \mathbf{Y}^{(m)} \mathbf{Y}'^{(m)}. \tag{3.10}$$

Again exchanging $\mathbf{Y}$ for $\mathbf{\Lambda}$ gives,

$$\mathbf{Q} = \frac{1}{M-1} \sum_{m=1}^{M} \mathbf{\Lambda}^{(m)} \mathbf{\Lambda}'^{(m)}, \tag{3.11}$$

where $\mathbf{\Sigma} = \mathbf{Q}/M$ (Equation 3.6).

Estimation of $\mathbf{\Sigma}$ allows for the calculation of the associated $\mathbf{P}$, which describes the correlations among $\mathbf{\Lambda}$, according to following quadratic,

$$\mathbf{P} = \mathbf{R}^{\frac{1}{2}} \mathbf{\Sigma} \mathbf{R}^{\frac{1}{2}}, \tag{3.12}$$

where $\mathbf{R}$ is the taken as diagonalised counterpart of $\mathbf{\Sigma}$ [13],

$$\mathbf{R}_{ij} = \frac{\delta_{ij}}{\mathbf{\Sigma}_{ij}}, \tag{3.13}$$

where δ is the Kronecker delta.

Dimensionality (N/M) has a notable effect on the reliability of the estimated $\mathbf{\Sigma}$. In high dimensional cases, the estimated $\mathbf{\Sigma}$ is prone to high levels of variation, as such in general $M \gg N$ must be satisfied for reliable estimation of $\mathbf{\Sigma}$ [39].

3.4 Selection of a Suitable S

Let us now define the S used in the new GOF procedure.

T^2 is a common choice for testing sample vectors,

$$T^2 = (\bar{\mathbf{X}} - \boldsymbol{\mu})\mathbf{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}), \tag{3.14}$$

where theoretically T^2 is distributed according to $\phi_{\chi^2}(S)$ with $\upsilon = N$. In practice however, this is often not achievable. In the present setting, when $M \gg N$ is not satisfied, $\mathbf{\Sigma}^{-1}$ is unstable. In cases such as these, the distribution of T^2 cannot be accurately described by $\phi_{\chi^2}(S)$ (Figure I).

This problem extends past just T^2, any S which considers the full Σ can fall victim to issues caused by dimensionality.

Hu et al. provide a suitable alternative, referred to as the Diagonalised T^2 (S) where $\mathbf{\Sigma}^{-1}$ is diagonalised along the leading diagonal [40],

$$S = \mathbf{\Lambda R \Lambda}', \tag{3.15}$$

where $\phi(S) \approx \phi_{\chi^2}(S)$. Note, that in systems where no correlations among the components of $\mathbf{\Lambda}$ exist, meaning when the components of $\mathbf{\Lambda}$ are i.i.d with $\sigma^2 = 1$, $\phi(S)$ will approach a $\phi_{\chi^2}(S)$.

3.5 The Characteristic Function

Now that both $\varrho(\mathbf{\Lambda}, \theta)$ and S have been defined, one can begin to derive $\phi(S)$ for use in the new GOF procedure.

$\phi(S)$ can be derived by integrating over all possible $\mathbf{\Lambda}$ for a given system [41],

$$\phi(S) = \langle \delta(S - \mathbf{\Lambda}'\mathbf{R}\mathbf{\Lambda}) \rangle = \int d\Lambda_1 ... d\Lambda_N \delta(S - \mathbf{\Lambda}'\mathbf{R}\mathbf{\Lambda}) \varrho(\mathbf{\Lambda}, \theta), \tag{3.16}$$

where here δ denotes the Dirac delta function. One can evaluate Equation 3.16 by using the integral representation of δ [41],

$$\phi(S) = \frac{1}{2\pi} \int d\mathbf{\Lambda}_1 ... d\mathbf{\Lambda}_N e^{-ikS} e^{ik\mathbf{\Lambda}'\mathbf{R}^*\mathbf{\Lambda}} \varrho(\mathbf{\Lambda}, \theta) dk = \frac{1}{2\pi} \int e^{-ikS} CF(k) dk \tag{3.17}$$

where k is a real-valued Fourier variable in the range $[-\infty, \infty]$.

Let us now derive the CF used in the evaluation of $\phi(S)$. From Equation 3.17, the CF can be defined as follows,

$$CF(k) = \int d\mathbf{\Lambda}_1 ... d\mathbf{\Lambda}_N e^{ik\mathbf{\Lambda}'\mathbf{R}^*\mathbf{\Lambda}} \varrho(\mathbf{\Lambda}, \theta). \tag{3.18}$$

By plugging Equation 3.8 into Equation 3.18, one arrives at,

$$CF(k) = \int \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Sigma}^*|^{\frac{1}{2}}} \left(e^{\left(ik\mathbf{\Lambda}'\mathbf{R}^*\mathbf{\Lambda} - \frac{1}{2}(\mathbf{\Lambda}-\boldsymbol{\mu})'\mathbf{\Sigma}^{*-1}(\mathbf{\Lambda}-\boldsymbol{\mu}) \right)} \right). \tag{3.19}$$

The multivariate Gaussian integral over all residuals in Equation 3.19 can be performed, see Mathai et al [36], leading to,

$$CF(k) = |\mathbf{I} - 2ik\mathbf{R}^*\mathbf{\Sigma}^*|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}\boldsymbol{\mu}'(\mathbf{I} - (\mathbf{I} - 2ik\mathbf{R}^*\mathbf{\Sigma}^*)^{-1})\mathbf{\Sigma}^{*-1}\boldsymbol{\mu} \right)}, \tag{3.20}$$

where $\mathbf{I}$ is the identity matrix. One can simplify the above expression, first by expanding,

$$CF(k) = |\mathbf{I} - 2ik\mathbf{R}^{*\frac{1}{2}}\mathbf{\Sigma}^*\mathbf{R}^{*\frac{1}{2}}|^{-\frac{1}{2}} e^{\left(ik\boldsymbol{\mu}'\mathbf{\Sigma}^{*-\frac{1}{2}}(\mathbf{R}^{*\frac{1}{2}}\mathbf{\Sigma}^*\mathbf{R}^{*-\frac{1}{2}})(\mathbf{I} - 2ik\mathbf{R}^{*\frac{1}{2}}\mathbf{\Sigma}^*\mathbf{R}^{*\frac{1}{2}})^{-1}\mathbf{\Sigma}^{*-\frac{1}{2}}\boldsymbol{\mu} \right)}, \tag{3.21}$$

and then by recalling $\mathbf{P}^* = \mathbf{R}^{*\frac{1}{2}}\mathbf{\Sigma}^*\mathbf{R}^{*\frac{1}{2}}$ (Equation 3.12), Equation 3.21 can be rewritten as follows,

$$CF(k) = \mid \mathbf{I} - 2ik\mathbf{P}^* \mid^{-\frac{1}{2}} e^{\left(ik\boldsymbol{\mu}'\mathbf{\Sigma}^{*-\frac{1}{2}}\mathbf{P}^*(\mathbf{I} - 2ik\mathbf{P}^*)^{-1}\mathbf{\Sigma}^{*-\frac{1}{2}}\boldsymbol{\mu} \right)}. \tag{3.22}$$

A well fitting model would lead to $\boldsymbol{\mu} \approx 0$, leaving only the leading term,

$$CF(k) = \mid \mathbf{I} - 2ik\mathbf{P}^* \mid^{-\frac{1}{2}} . \tag{3.23}$$

In order to have the CF represented in a more usable format, Equation 3.22 can be deformed from matrix notation and represented by its eigenvalues (λ) [36]. First, one can define an orthogonal matrix, $\mathbf{A}$, such that $\mathbf{AIA^{-1}} = \mathbf{I}$ and $\mathbf{A^{-1}P^*A} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix containing the λ of $\mathbf{P}^*$. Then, noting that both $|\mathbf{I}|$ and $|\mathbf{D}|$ are equal to the product of their diagonal entries [36], one has,

$$CF(k) = |\mathbf{AIA^{-1}} - 2ik\mathbf{A^{-1}P^*A}|^{-\frac{1}{2}} = |\mathbf{I} - 2ik\mathbf{D}|^{-\frac{1}{2}} = \prod_{i=1}^{N} (1 - 2ik\lambda_i)^{-\frac{1}{2}}, \tag{3.24}$$

where λ are the eigenvalues of $\mathbf{P}$. Note, if one were to go through the above sequence in the in the case that the correlations among the components of Λ where uncorrelated, one would expect that $\lambda = 1$ for all λ. In this case the resulting CF is that associated with $\phi_{\chi^2}(S)$, where if $\lambda = 1$,

$$CF(k) = \prod_{i=1}^{N}(1 - 2ik\lambda_i)^{-\frac{1}{2}} = (1 - 2ik)^{-\frac{N}{2}}. \tag{3.25}$$

Applying the inverse Fourier transform to Equation 3.24 returns Equation 3.17 and yields the required $\phi(S)$,

$$\phi(S) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(-ikS)} \prod_{i=1}^{N}(1 - 2ik\lambda_i)^{-\frac{1}{2}} dk = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{(-ikS)} CF(k) dk. \tag{3.26}$$

The integral presented in Equation 3.26 provides a suitable $\phi(S)$ for use in the new GOF procedure. Moving forward, simulated estimates of $\mathbf{\Sigma}^*$, $\mathbf{R}^*$ and $\mathbf{P}^*$ are used in the evaluation of $\phi(S)$.

3.6 The $\boldsymbol{\mu}$ and σ^2 of $\phi(S)$

Following the deviation of $\phi(S)$, one can now look at both its $\boldsymbol{\mu}(\langle S\rangle)$ and $\sigma^2(\langle S^2\rangle - \langle S\rangle^2)$ analytically, and draw a comparison to $\phi_{\chi^2}(S)$.

Formally, the CF is the expected value of the exponent presented in Equation 3.26 [36],

$$CF(k) = \langle e^{ikS}\rangle. \tag{3.27}$$

Taylor expansion of Equation 3.27 allows for the identification of $\langle S\rangle$ and $\langle S^2\rangle$,

$$CF(k) = 1 + ik\langle S\rangle - \frac{k^2}{2}\langle S^2\rangle, \tag{3.28}$$

where $\langle S\rangle$ and $\langle S^2\rangle$ are the first and second moments respectively.

Taylor expanding the CF presented in Equation 3.26 allows for the determination of $\langle S\rangle$ and $\langle S^2\rangle$ in terms of the λ of P,

$$CF(k) = \prod_{i=1}^{N} 1 + ik\lambda_i - \frac{3k^2}{2}\lambda_i^2, \tag{3.29}$$

where by expanding out the product,

$$CF(k) = (1 + ik\lambda_1 - \frac{3k^2}{2}\lambda_1^2)(1 + ik\lambda_2 - \frac{3k^2}{2}\lambda_2^2)...(1 + ik\lambda_N - \frac{3k^2}{2}\lambda_N^2). \tag{3.30}$$

Further expansion leads one to the following, where $\langle S \rangle$ and $\langle S^2 \rangle$ are readily identifiable,

$$CF(k) = 1 + ik \sum_{i=1}^{N} \lambda_i - k^2 (\sum_{1 \leq i < j \leq N}^{N} \lambda_i \lambda_j) - \frac{3k^2}{2} \sum_{i=1}^{N} \lambda_i^2, \tag{3.31}$$

where terms above $O(k^2)$ have been removed.

Equation 3.31 allows for the identification of $\langle S \rangle$ as the sum of λ,

$$\langle S \rangle = \sum_{i=1}^{N} \lambda_i = tr(\mathbf{P}) = N. \tag{3.32}$$

Comparing with the χ^2-GOF procedure, one finds that both $\phi(S)$ and $\phi_{\chi^2}(S)$ have the same $\boldsymbol{\mu}$, where $\phi_{\chi^2}(S)$ has $\boldsymbol{\mu} = \upsilon = N$ [20].

Equation 3.31 also allows for σ^2 to be determined, where σ^2 is equal to the second moment minus $\boldsymbol{\mu}^2$,

$$\sigma^2 = \langle S^2 \rangle - \langle S \rangle^2, \tag{3.33}$$

$$\sigma^2 = 3 \sum_{i=1}^{N} \lambda_i^2 + 2 (\sum_{1 \leq i < j \leq N}^{N} \lambda_i \lambda_j) - (\sum_{i=1}^{N} \lambda_i)^2 = 2 \sum_{i=1}^{N} \lambda_i^2. \tag{3.34}$$

Again comparing with $\phi_{\chi^2}(S)$, where $\sigma^2 = 2\upsilon$ [20], it is not possible to say $\sigma^2(\phi(S)) = \sigma^2(\phi_{\chi^2}(S))$ is always true. Indeed, if $N = 1$, then the CF would describe the $\phi_{\chi^2}(S)$, with $\boldsymbol{\mu} = \sigma^2 = 1$, but this is cannot be said definitively for an arbitrary choice of N.

Solving for the minimum $\sigma^2(\phi(S))$ gives further insight into whether $\sigma^2(\phi(S)) = \sigma^2(\phi_{\chi^2}(S))$ for $N > 1$. On that note, let us set $N = 2$ and evaluate $\sigma^2(\phi(S))$ and $\boldsymbol{\mu}(\phi(S))$,

$$\boldsymbol{\mu}(\phi(S)) = \sum_{i=1}^{2} \lambda_i = \lambda_1 + \lambda_2 = 2, \tag{3.35}$$

$$\sigma^2(\phi(S)) = 2 \sum_{i=1}^{2} \lambda_i^2 = 2\lambda_1 + 2\lambda_2 = 2\lambda_1^2 + 2(2 - \lambda_1)^2. \tag{3.36}$$

Solving $\frac{d\sigma^2}{d\lambda_1} = 0$, allows for the smallest values of λ_1 to be found,

$$\frac{d\sigma^2}{d\lambda_1} = 4\lambda_1 + 4(\lambda_1 - 2) = 0, \tag{3.37}$$

thus $\lambda_1 = 1$. As $\lambda_1 + \lambda_2 = 2$ (Equation 3.32), it can be said $\lambda_1 = \lambda_2 = 1$, meaning $\sigma^2 = 4 = 2N$. Thus, $\sigma^2(\phi(S)) = 2N$ is the minimum value $\sigma^2(\phi(S))$ can take, which is the same as $\sigma^2(\phi_{\chi^2}(S))$.

If the assumption made by the χ^2-GOF procedure that correlations among the components of $\boldsymbol{\Lambda}$ are i.i.d is invalid, then one would expect the evaluated $\sigma^2(\phi(S))$ to differ from $\sigma^2(\phi_{\chi^2}(S))$. Since $\sigma^2(\phi(S))$ can only increase relative to $\sigma^2(\phi_{\chi^2}(S))$, one would expect an increased σ^2 when correlations among the components of $\boldsymbol{\Lambda}$ are non i.i.d.

3.7 The Marchenko-Pastur Law

This section defines the Marchenko-Pastur (MP) law (Appendix C).

The MP law describes a typical distribution of λ for a given random matrix made up of the product of i.i.d sample vectors, with $\boldsymbol{\mu} = 0$ and $\sigma^2 = 1$ [42]. Let us define,

$$\zeta = \frac{1}{N}\mathbf{Y}\mathbf{Y}^T, \tag{3.38}$$

where as before $\mathbf{Y}$ is a sample vector of N length, with i.i.d entries. The MP law then describes the distribution of the λ of ζ.

In the case that the correlations among the components of $\mathbf{\Lambda}$ are i.i.d, the MP law provides a typical distribution of λ for $\mathbf{P}$ ($\psi(\lambda)$) [42],

$$\psi(\lambda) = \begin{cases} \dfrac{1}{2\pi\lambda\sqrt{\frac{N}{M}}}\sqrt{(b-\lambda)(\lambda-a)} & a < \lambda < b \\ 0 & otherwise \end{cases}, \tag{3.39}$$

where a and b are defined as follows,

$$a = (1 - \sqrt{\frac{N}{M}})^2, \tag{3.40}$$

$$b = (1 + \sqrt{\frac{N}{M}})^2, \tag{3.41}$$

where a and b also define the lower and upper bounds of $\psi(\lambda)$ respectively. Here, when $M \longrightarrow 1$, one would expect $\lambda = 1$ for all λ. In this case, evaluation of the CF presented in Equation 3.25 would result in $\phi_{\chi^2}(S)$, with $\boldsymbol{\mu} = \sigma^2 = 1$.

4 Methods

This section details the Gil-Pelaez theorem, its application to the evaluation of $\phi(S)$ and the cumulative distribution function ($F(S)$), followed by integration methods and the contents of the estimate ensembles averages generated by the prototype models.

4.1 The Gil-Pelaez Theorem

The Gil-Pelaez theorem allows for the evaluation of the integral presented in Equation 3.26. Gil-Pelaez demonstrates explicitly that if a given CF is of the form,

$$CF(k) = \int_{-\infty}^{\infty} e^{(ikS)}\phi(S)dS, \tag{4.42}$$

then its associated $\phi(S)$ can be evaluated by applying the inversion theorem [21],

$$\phi(S) = \frac{1}{\pi} \int_0^\infty \mathcal{R}\{e^{(-ikS)} CF(k)\} dk, \tag{4.43}$$

where $\mathcal{R}\{...\}$ refers to the real components of a given function.

On evaluating the integral presented in Equation 4.43, Imhof and Davies provide examples of the use of numerical methods [22, 23]. This thesis takes a similar approach, using the trapezoidal rule to approximate the integral presented.

4.2 Evaluation of $\phi(S)$

Application of the trapezoidal rule allows for the approximation of Equation 4.43 [18],

$$\phi(S) = \frac{dk}{\pi} \left(\frac{1}{2} + \sum_{j=1}^{n} w_j \mathcal{R}\left(e^{ik_j S} CF(k_j)\right) \right), \tag{4.44}$$

where n is the number of points evaluated, w the quadrature weights, k is a real valued Fourier variable in the range $[-\infty, \infty]$, and,

$$dk = \frac{2\pi}{U - L}, \tag{4.45}$$

where U and L define the upper and lower range of $\phi(S)$ respectively. For information on the leading 1/2 and a full derivation of Equation 4.44, see the Appendix (Appendix A).

The six-sigma rule can be used to select U and L [26], such that U - L spans a range where all possible S for a given system could fall. Using this approach,

$$U = \sum_{i=0}^{N} \lambda_i + \sum_{i=0}^{N} \lambda_i^2, \tag{4.46}$$

and $L = 0$ by definition as $S > 0$ for all S.

In order to improve the reliability of the evaluated $\phi(S)$, Equation 4.44 is repeated a number of times, with an increasing number of points per iteration. Romberg's method is then applied [18]. Romberg's method is an iterative process, in which increasingly accurate evaluations of a given integral can be used to generate a triangular array of increasingly accurate estimates of said integral. Romberg's method then stops when the results along the diagonal of the triangular array begin to converge, demonstrating that the integral has been approximated accordingly. Romberg's method ensures that a suitable amount of points have been used to approximate an integral evaluated from 0 to ∞, as in Equation 4.44. For further detail the reader is directed to Press et al [18].

4.3 Evaluation of $F(S)$

This sections presents the method used to evaluate the cumulative distribution function $F(S)$. $F(S)$ presents the probability that S takes a value less than or equal to a specified value in evaluated space,

$$F(S) = \int_0^S \phi(S) dS. \tag{4.47}$$

In the evaluation of $F(S)$, the Gil-Pelaez theorem is again applied. Imhof demonstrates that any given F, with a CF of the form shown in Equation 4.42, can be evaluated using the Gil-Pelaez theorem [21],

$$F(S) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty k^{-1} \mathcal{I}\{e^{(-ikS)} CF(k)\} dk, \tag{4.48}$$

where $\mathcal{I}\{...\}$ refers to the imaginary components of a function. The trapezoidal rule can again be applied,

$$F(S) = \frac{1}{2} - \frac{dk}{\pi} \left(\frac{N-k}{2} + \sum_{j=0}^{n} w_j \mathcal{I} \left(\frac{e^{ik_j S} CF(k_j)}{k_j} \right) \right). \tag{4.49}$$

For information on the leading term, and a full derivation of Equation 4.49, the reader is again directed to the Appendix (Appendix A). In the case of Equation 4.49, Romberg's method is again applied to ensure suitable approximate of an infinite integral.

4.4 Prototype Model Trajectories

This section describes the contents of the estimate ensemble averages generated via the three prototype models.

The observable used in the estimate ensemble averages, unless otherwise stated, is the squared displacement (SD),

$$y^{(m)}(t) = \mid y^{(m)}(t) - y^{(m)}(0) \mid^2, \tag{4.50}$$

where $y^{(m)}(t)$ denotes the SD of a given trajectory at time, t, with $t = (1, ..., N)$ [9]. The ensemble at a given t is then the average position of the sample population at each sampling point,

$$\langle y(t) \rangle = \frac{1}{M} \sum_{m=1}^{M} y^{(m)}(t), \tag{4.51}$$

where $\langle ... \rangle$ denotes the ensemble average of a given observable.

The choice of SD as the time-dependant observable is due to both its commonality in literature [2, 4, 5, 8, 9, 10, 11, 12, 13], and ease of calculation in all selected prototype models (Equation 4.50).

5 Results and Discussion

This section presents the results of a series of tests validating the new GOF procedure. Each result is preceded by a description of the testing procedure and followed by a discussion of the outcome.

There are four different $\phi(S)$ used in the testing of the new GOF procedure. They are as follows:

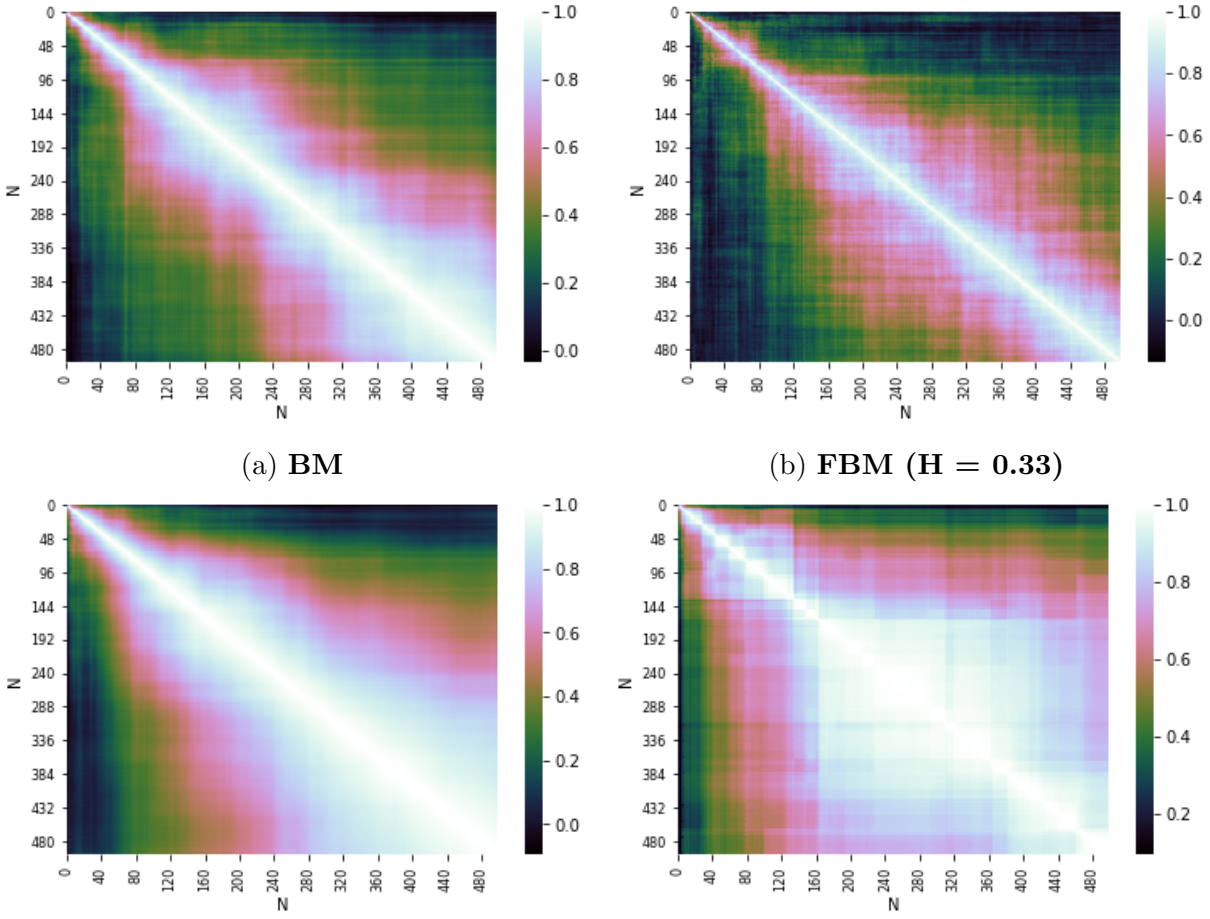
- $\phi(S)$ - $\mathbf{\Sigma}$ is estimated via Equation 3.6, $\phi(S)$ is then evaluated using the methods presented in Section 4.2.
- $\phi(S)_{true}$ - A histogram of $z = 1000$ calculated S is plotted, serving as the 'true' $\phi(S)$ for a given system, where is S calculated using Equation 3.15, with $\mathbf{\Sigma}$ estimated from Equation 3.6.
- $\phi(S)^*$ - $\mathbf{\Sigma}^*$ is derived analytically for a specific system, $\phi(S)$ is then evaluated using the methods presented in Section 4.2. $\phi(S)$ in this case is the analytically exact $\phi(S)$, $\phi(S)^*$.
- $\phi(S)^*_{true}$ - A histogram of $z = 1000$ calculated S is plotted, where is S calculated using $\mathbf{\Sigma}^*$. This histogram then serves as the 'true' and analytically exact $\phi(S)$ for a given system.
- $\phi_{\chi^2}(S)$ - The χ^2-distribution representing S for a given system.

5.1 Demonstrating Correlations Among the Components of $\mathbf{\Lambda}$

This section sets out to demonstrate the existence of correlations among the components of $\mathbf{\Lambda}$, providing evidence that assumption made by the χ^2-GOF procedure, in which the correlations among the components of $\mathbf{\Lambda}$ are said to be i.i.d, is invalid.

If the assumption made by the χ^2-GOF procedure was correct, one would expect $\mathbf{P}$ to be a $N \times N$ random matrix, with no visible correlation pattern among the components of $\mathbf{\Lambda}$.

The prototype models are used to generate estimate ensembles made up of M trajectories over N sampling points (Section 4.4), a model is then fit to each (Figure B) using the WLS-ICE method (Appendix E) and the residuals calculated (Equation 3.7). Visualisation of the correlations among the components of $\mathbf{\Lambda}$ can be achieved by plotting the $\mathbf{P}$ matrix. The $\mathbf{P}$ matrix, by definition, defines the correlations among the components $\mathbf{\Lambda}$ and is calculated as in Equation 3.12. Visualisation of $\mathbf{P}$ will allow for evaluation of the correlation type of each system, determining whether, in the case of the four prototype models, the correlation types are indeed, i.i.d, or otherwise as expected. The results are shown in Figure 3.



(a) **BM** (b) **FBM (H = 0.33)**

(c) **FBM (H = 0.66)** (d) **CTRW**

Figure 3: **Typical Correlation Conditions Among Prototype Models**. Figures 3 (a) - (d) present the typical correlation conditions associated with the BM ($M = 250, N = 500$), FBM ($H = 0.33, M = 250, N = 500$), FBM ($H = 0.66, M = 250, N = 500$) and CTRW ($M = 250, N = 500$) prototype models respectively. Note, in all cases the off-diagonal components of $\mathbf{P}$ are non-zero.

In the four testing environments plotted, the correlations among $\mathbf{\Lambda}$ are non i.i.d (Figure 3). It can be observed in all cases that the off-diagonal components of $\mathbf{P}$ are non-zero, meaning there are correlations among the components of $\mathbf{\Lambda}$. Thus, it can be said that the assumption taken by the χ^2-GOF procedure that correlations among the components of $\mathbf{\Lambda}$ are i.i.d, is invalid.

5.2 Validation of the Inversion Method and Proposed $\phi(S)$

This section presents proof-of-concept for the new GOF procedure. The method for generating $\phi(S)$ via numerical inversion of the characteristic function (Section 3.5) is validated, as is the ability of the evaluated $\phi(S)$ to correctly represent the distribution of S.

In the case of the BM prototype model, one can derive $\mathbf{\Sigma}^*$ analytically (Appendix B). Removal of the random effects of $\mathbf{\Sigma}$ estimation allows for only the inversion method (Section 4.2), and the ability of the evaluated $\phi(S)^*$ to match $\phi(S)^*_{true}$ to be tested. A close to exact match between $\phi(S)^*$ and $\phi(S)^*_{true}$ will hence demonstrate the feasibility of the methods used in the new GOF procedure.

The BM prototype model, with $M = 1000$ trajectories and $N = 100$ sampling points, is used to generate $\phi(S)^*$ and $\phi(S)^*_{true}$. The quality of match between $\phi(S)^*$ and $\phi(S)^*_{true}$ is then evaluated. The result of the comparison between $\phi(S)^*$ and $\phi(S)^*_{true}$ is presented in Figure 4.

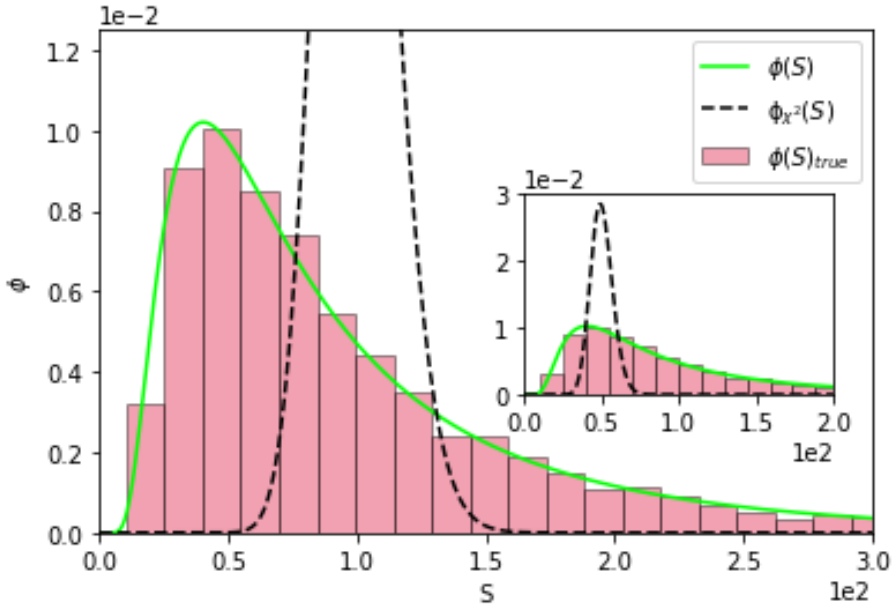


Figure 4: **Validation of the Inversion Method and Proposed $\phi(S)$ - $\phi(S)^*$ vs. $\phi(S)^*_{true}$.** Figure 4 presents a comparison between $\phi(S)^*$ vs. $\phi(S)^*_{true}$. The test ensembles were generated using the BM prototype model, with $M = 1000$ trajectories and $N = 100$ sampling points. In comparing $\phi(S)$ and $\phi_{\chi^2}(S)$, notice $\phi(S)$ has an asymmetric and broader distribution when compared to $\phi_{\chi^2}(S)$, and is a closer match to $\phi(S)_{true}$.

Whenever random contributions from the estimation of $\mathbf{\Sigma}$ are removed, $\phi(S)^*$ is a near exact match to $\phi(S)^*_{true}$ (Figure 4). This validates the selected inversion method and demonstrates proof-of-concept for the new GOF procedure.

$\phi(S)^*$ provides an improved prediction of $\phi(S)^*_{true}$ when compared to $\phi_{\chi^2}(S)$, with both distributions centered around N, but the asymmetric distribution of $\phi(S)^*$ proving to be a better prediction of $\phi(S)^*_{true}$ (Figure 4).

The asymmetry and increased broadness of $\phi(S)^*$ can be described by the increased σ^2 of $\phi(S)^*_{true}$ relative to that of $\phi_{\chi^2}(S)$ (Section 3.6). The correlations among the components of $\mathbf{\Lambda}$ cause $\sigma^2(\phi(S)^*)$ to deviate from the theoretical, with values above the minimum $\sigma^2(\phi(S)^*) = 2N$. The inability of the χ^2-GOF procedure to account for changes in σ^2 results in a poorer match , with $\phi_{\chi^2}(S)$ proving to be a poor predictor of $\phi(S)^*_{true}$.

5.3 Comparison of the $\boldsymbol{\mu}$ and σ^2 of $\phi(S)$ and $\phi_{\chi^2}(S)$

This section aims to numerically demonstrate the driving factor causing the difference between $\phi(S)^*$ and $\phi_{\chi^2}(S)$ (Figure 4).

It has been demonstrated that $\phi(S)$ and $\phi_{\chi^2}(S)$ share the same $\boldsymbol{\mu}$ (Section 3.6), though when one looks at the σ^2 of both ϕ, it can be shown that $\sigma^2(\phi(S))$ is likely always greater then $\sigma^2(\phi_{\chi^2}(S))$ once $N > 1$ (Section 3.6). This can also be demonstrated numerically.

The BM prototype model is used to demonstrate how $\sigma^2(\phi(S))$ (Equation 3.34) changes relative to $\sigma^2(\phi_{\chi^2}(S))$ as N is increased from $N = 2$ to $N = 50$ sampling points, where $\sigma^2(\phi_{\chi^2}(S)) = 2N$. This is done at $M = 10$, $M = 100$ and $M = 1000$ trajectory counts. The result can be found in Figure 5.

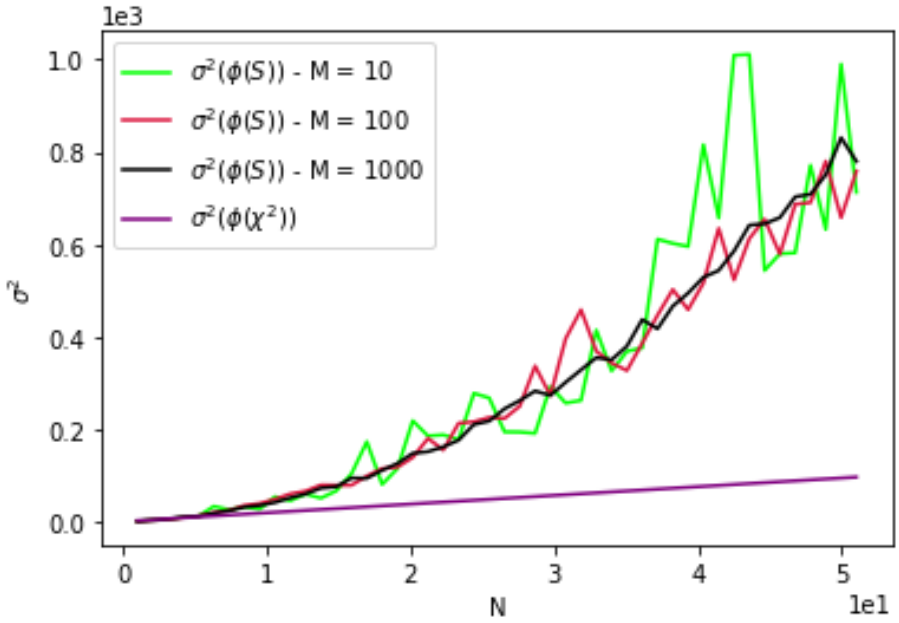


Figure 5: **Comparison of the μ and σ^2 of $\phi(S)$ and $\phi_{\chi^2}(S)$ - $\sigma^2(\phi(S))$ and $\sigma^2(\phi_{\chi^2}(S))$ vs. N**. Figure 5 shows $\sigma^2(\phi(S))$ and $\sigma^2(\phi_{\chi^2}(S))$ as a function of N. The BM prototype model was used to generate all estimate ensemble averages, with $M = 10$, $M = 100$ and $M = 1000$ trajectories and $N = 50$ sampling points. Notice, in all cases, once $N > 1$, one can observe increasing deviations between $\sigma^2(\phi(S))$ and $\sigma^2(\phi_{\chi^2}(S))$.

As N increases, one can see an exponentially increasing deviation between $\sigma^2(\phi(S))$ and $\sigma^2(\phi_{\chi^2}(S))$ (Figure 5). The increased $\sigma^2(\phi(S))$ results in a broader asymmetric $\phi(S)$ when compared to $\phi_{\chi^2}(S)$, and indeed can be considered to be the primary driver of difference between $\phi(S)$ and $\phi_{\chi^2}(S)$.

5.4 Investigation of the λ of P

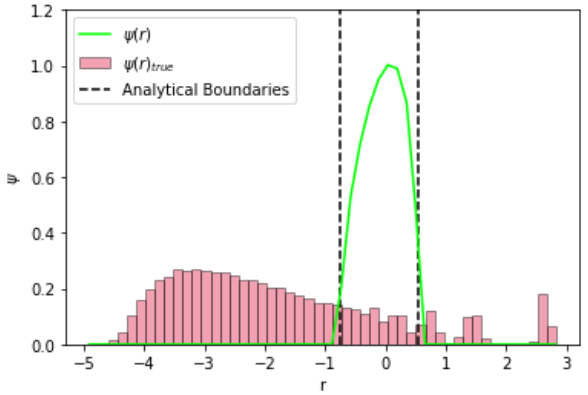
This section further investigates the effects of increased σ^2 on $\phi(S)$, in terms the of the eigenvalues (λ) of the Pearson correlation matrix ($\mathbf{P}$) (Equation 3.12).

Recall, that if one where to assume correlations among $\mathbf{\Lambda}$ where i.i.d, as is assumed in the χ^2-GOF procedure, one could consider P to be a $N \times N$ random matrix (Section 5.1). In this case one would expect the distribution of λ to tend towards the Marchenko-Pastur (MP) law (Section 3.7), where the MP law proposes a typical distribution of λ ($\psi(\lambda)$) for a given random matrix. The MP law proposes not only a typical $\psi(\lambda)$, but also defines an upper and lower boundary to λ, determined by $1+\sqrt{N/M})^2$ and $1-\sqrt{N/M})^2$ respectively (Section 3.7).

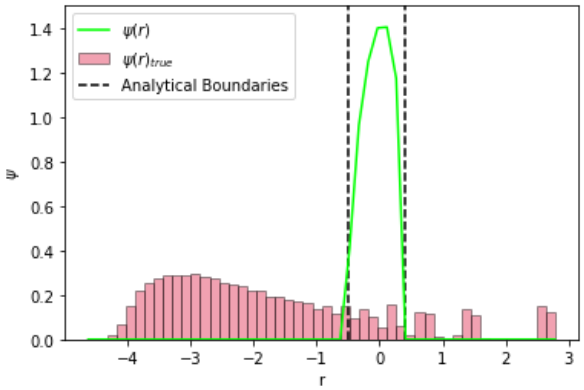
If the correlations among $\mathbf{\Lambda}$ where i.i.d, one would expect $\psi(\lambda)$, determined by the MP law,

to accurately predict the λ distribution of P. In the optimal case, as $M \longrightarrow \infty$, $\sqrt{N/M}$ becomes infinitesimal, meaning one would expect $\lambda = 1$ for all λ. In this case, inversion of the associated CF would produce $\phi_{\chi^2}(S)$ (Section 3.7).

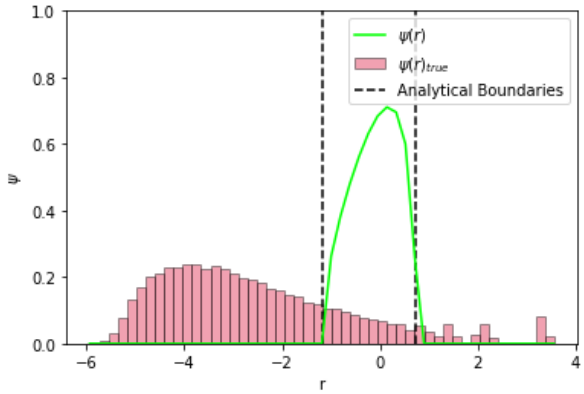
Figure 6 presents a comparison between $\psi(r)_{true}$, where $r = \ln \lambda$, and $\psi(r)$, where $\psi(r)$ is the distribution of r proposed by the MP law. Here, $\psi(r)_{true}$ has been evaluated through $z = 1000$ repeated evaluations of $\mathbf{P}$ generated from BM prototype model estimate ensembles of a fixed number of M trajectories and N sampling points, and the associated r plotted on a histogram, serving as a measure of the 'true' distribution of r. The choice to use log spectrum was purely for readability purposes, see the appendix for further details regarding the MP law and $\psi(r)$ (Appendix C).



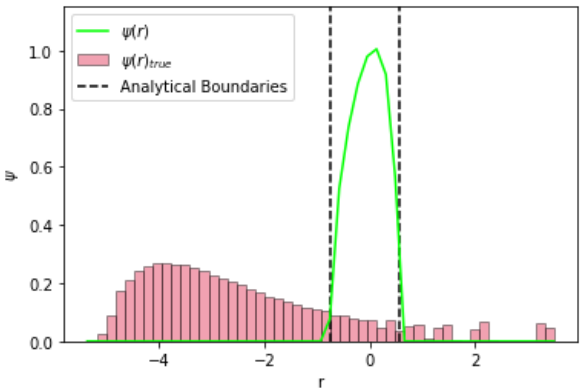
(a) **M = 250, N = 25**



(b) **M = 500, N = 25**



(c) **M = 250, N = 50**



(d) **M = 500, N = 50**

Figure 6: **Investigation of the λ of P - $\psi(r)_{true}$ vs. $\psi(r)$.** Figures 6 (a) - (d) presents a comparison between $\psi(r)_{true}$ and $\psi(r)$ under different M and N settings. The test ensembles were generated using the BM prototype model, with M and N defined in the relevant subheading. Notice, in all cases one can see a much broader, asymmetric, distribution of r than is predicted by $\psi(r)$.

In the tested scenarios, it is clear that $\psi(r)_{true}$ is not well approximated by $\psi(r)$ (Figure 6). In all cases, one can see a much broader $\psi(r)_{true}$ than predicted by $\psi(r)$, with only a small portion of r contained within the analytical boundaries determined by $1+\sqrt{N/M})^2$ and $1-\sqrt{N/M})^2$.

The broadness of $\psi(r)_{true}$ is a indicator of the increased σ^2 in $\psi(r)$. This increased σ^2 results in the observed difference between $\phi(S)$ and $\phi_{\chi^2}(S)$ (Figure 4).

5.5 The Evaluation of $\phi(S)$ in Varied Correlation Conditions

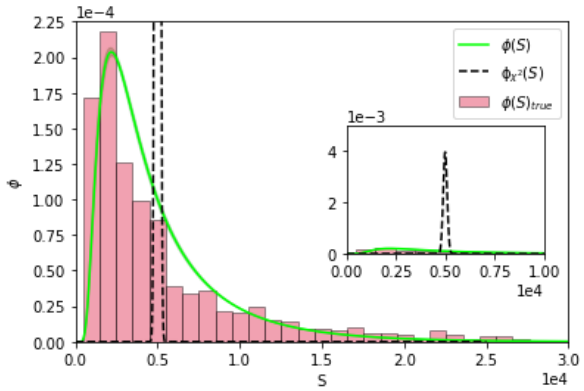
This section presents the testing of the new GOF procedure under varied correlation types, using large scale ensembles generated from all three prototype models. Here, correlation type defines different types of Pearson correlation matrix ($\mathbf{P}$) (Figure 3), where $\mathbf{P}$ defines the correlation among the components of $\mathbf{\Lambda}$ (Section 5.1).

If the new GOF procedure is valid regardless of the type of correlation among the components of $\mathbf{\Lambda}$, one would expect a valid match between $\phi(S)$ and $\phi(S)_{true}$ in all testing environments.

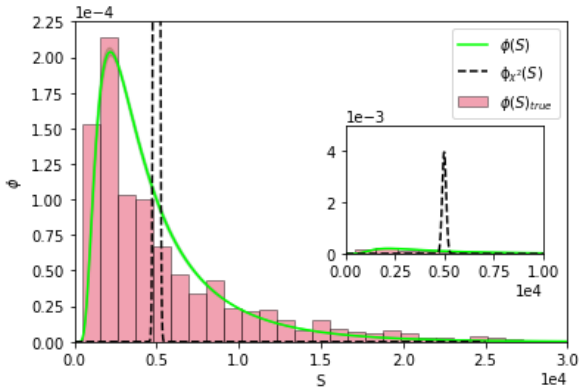
The BM, FBM ($H = 0.33, H = 0.66$) and the CTRW prototype models are used to generate a series of estimate ensemble averages, with each model producing a estimate ensemble average with different correlations among the components of $\mathbf{\Lambda}$ (Section 5.1), which can then be used to generate suitable estimates of $\mathbf{\Sigma}$ for use in the evaluation of $\phi(S)$ (Sections 3.5, 4.2). $\phi(S)$ and $\phi(S)_{true}$ are then compared for all four testing environments. In all testing environments, M was selected to be large enough that the multivariate CLT is valid (Section 3.2).

5.5.1 BM Prototype Model

The section presents the testing of the new GOF procedure using the BM prototype model. The results can be found in Figure 7.



(a) **M = 250, N = 5000**



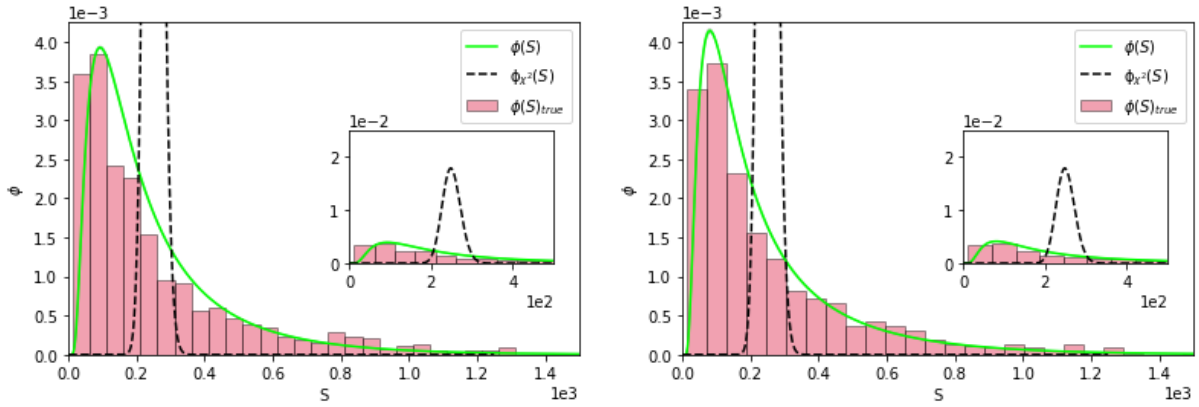
(b) **M = 500, N = 5000**

Figure 7: **The Evaluation of $\phi(S)$ in the BM Prototype Model - $\phi(S)$ vs. $\phi(S)_{true}$.** Figures 7 (a) and (b) present the comparisons between $\phi(S)$ and $\phi(S)_{true}$ at varied M using the BM prototype model. Here, $\phi(S)$ is an excellent predictor of $\phi(S)_{true}$, with an almost exact match between $\phi(S)$ and $\phi(S)_{true}$ in both cases.

In the BM prototype model, given M is large enough, the new GOF procedure evaluates a $\phi(S)$ that is a close-to-exact match to $\phi(S)_{true}$ (Figure 7). In both settings, $\phi(S)$ matches $\phi(S)$ in terms of shape and numerical value in both the lower and tail ends of the distribution. Increasing M has a minimal effect on the match quality between $\phi(S)$ and $\phi(S)_{true}$, as M is large enough for the multivariate CLT to be satisfied. Thus, when $\mathbf{P}$ is of the typical form for BM (Figure 3 (a)), the new GOF procedure can evaluate a reliable $\phi(S)$.

5.5.2 FBM (H = 0.66)

The section presents the testing of the new GOF procedure using the FBM prototype model with $H = 0.66$. The results can be found in Figure 8.



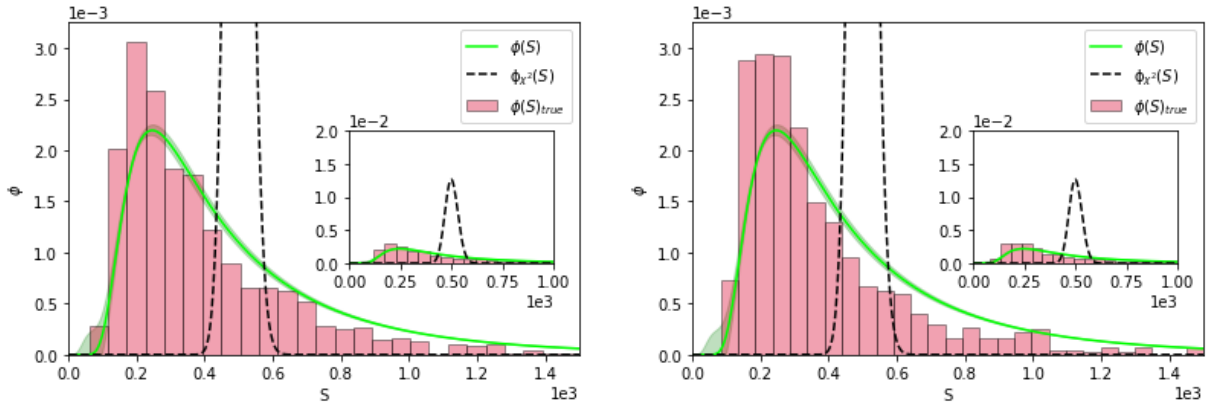
(a) **M = 250, N = 250** (b) **M = 500, N = 250**

Figure 8: **The Evaluation of $\phi(S)$ in the FBM ($H = 0.66$) Prototype Model - $\phi(S)$ vs. $\phi(S)_{true}$.** Figures 8 (a) and (b) present the comparisons between $\phi(S)$ and $\phi(S)_{true}$ at varied M using the FBM ($H = 0.66$) prototype model. One can again observe a close match between $\phi(S)$ and $\phi(S)_{true}$ in all the tested cases.

In the FBM ($H = 0.66$), given M is large enough, the new GOF procedure evaluates a $\phi(S)$ that is a close-to-exact match to $\phi(S)_{true}$ (Figure 8). In both settings, $\phi(S)$ matches $\phi(S)$ in terms of shape and numerical value in both the lower and tail ends of the distribution, particularly in the $M = 250$ and $N = 250$ case (Figure 8 (a)), with $\phi(S)$ being essentially an exact match to $\phi(S)_{true}$. Again, increasing M has a minimal effect on the match quality between $\phi(S)$ and $\phi(S)_{true}$. In conditions where correlations among $\mathbf{\Lambda}$ produce a $\mathbf{P}$ as seen in Figure 3 (c), the evaluated $\phi(S)$ is a valid predictor of $\phi(S)_{true}$.

5.5.3 FBM (H = 0.33)

The section presents the testing of the new GOF procedure using the FBM prototype model with $H = 0.33$. The results can be found in Figure 9.



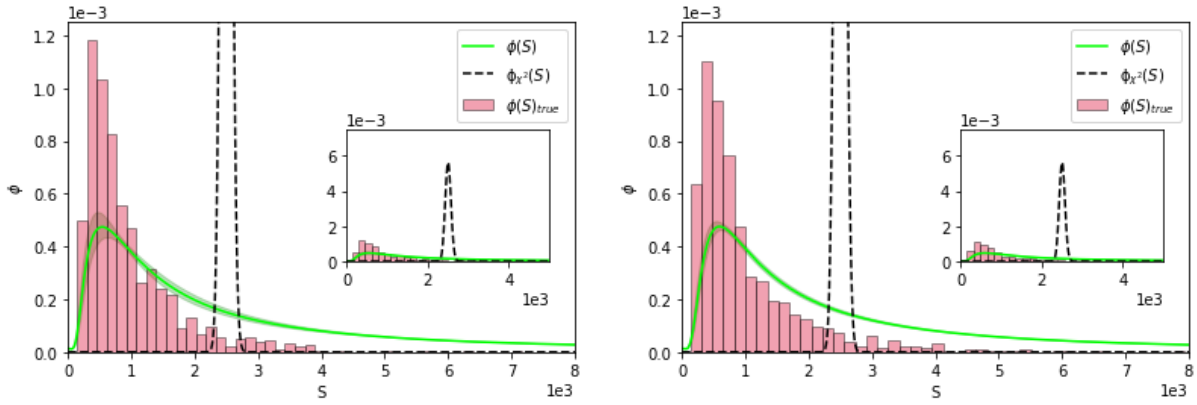
(a) **M = 250, N = 500** (b) **M = 500, N = 500**

Figure 9: **The Evaluation of $\phi(S)$ in the FBM ($H = 0.33$) Prototype Model - $\phi(S)$ vs. $\phi(S)_{true}$.** Figures 9 (a) and (b) present the comparisons between $\phi(S)$ and $\phi(S)_{true}$ at varied M using the FBM ($H = 0.33$) prototype model. Notice in this case, $\phi(S)$ is not as accurate a predictor of $\phi(S)_{true}$ relative to BM (Figure 7) and FBM ($H = 0.66$) (Figure 8) cases.

In the FBM ($H = 0.33$), given M is large enough, the new GOF procedure evaluates a $\phi(S)$ that is a valid match to $\phi(S)_{true}$ (Figure 9). In comparison with the BM (Figure 7) and FBM ($H = 0.66$) (Figure 8) cases, the matches between $\phi(S)$ and $\phi(S)_{true}$ when **P** is of the form shown in Figure 3 (b), are in general poorer. In the tested cases above, the majority of $\phi(S)_{true}$ is covered by $\phi(S)$, and the tail end is well estimated, but there is no close match around the peak of $\phi(S)_{true}$. Increasing M again has a minimal effect as M is large enough for the multivariate CLT to be satisfied.

5.5.4 CTRW

The section presents the testing of the new GOF procedure using the CTRW prototype model. The results can be found in Figure 10.



(a) **M = 250, N = 2500** (b) **M = 500, N = 2500**

Figure 10: **The Evaluation of $\phi(S)$ in the CTRW Prototype Model - $\phi(S)$ vs. $\phi(S)_{true}$.** Figures 10 (a) and (b) present the comparisons between $\phi(S)$ and $\phi(S)_{true}$ under super-linear correlation conditions at varied M using the CTRW prototype model. Here, notice although $\phi(S)$ covers most of $\phi(S)_{true}$, it fails to accurately predict the peak and tail.

Within the CTRW prototype model, given M is large enough, the new GOF procedure evaluates a $\phi(S)$ that is a valid, but not exact, match to $\phi(S)_{true}$ (Figure 10). In cases where the calculated **P** is of the form as in Figure 3 (d), $\phi(S)$ tends to cover the majority of $\phi(S)_{true}$, but fails to correctly predict the peak and tail end of $\phi(S)$. Increasing M has no effect in this case, with no improvement in match quality as M is increased.

5.5.5 Discussion

The new GOF procedure is applicable in varying correlation conditions (various forms of **P**), producing a suitable $\phi(S)$ in all testing environments (Figures 7, 8, 9, 10). In the BM prototype model, $\phi(S)$ produces a close-to-exact match to $\phi(S)_{true}$ in all tested cases (Figure 7). In the case of both FBM prototype models, $\phi(S)$ is almost exact match to $\phi(S)_{true}$ when $H = 0.66$ (Figures 8), and a valid match when $H = 0.33$ (Figures 9). In the case of the CTRW (Figure 11), $\phi(S)$ is a reasonable match to $\phi(S)_{true}$, but is not a great predictor of the peak and tail end of $\phi(S)_{true}$.

From the testing scenarios presented (Figures 7, 8, 9, 10), it is clear that the form of P plays a role in quality of match between $\phi(S)$ and $\phi(S)_{true}$. It is clear from the almost exact match between $\phi(S)$ and $\phi(S)_{true}$, that the typical form of **P** for both the BM and FBM ($H = 0.66$) provide optimal conditions for the new GOF procedure. In the case that the off-diagonals of **P** are as in Figures 3 (b) and (d), the difference between the evaluated $\phi(S)$ and $\phi(S)_{true}$ increases. One could perhaps say that when **P** is of the form shown in Figures 3 (b) and (d), a much larger M in needed to for the multivariate CLT to be valid, but this can't be said definitively.

Overall, given M is large enough for the multivariate CLT to be valid, the new GOF procedure can produce a reliable $\phi(S)$ regardless of the form of $\mathbf{P}$, and provides a more accurate representation of $\phi(S)_{true}$ than the traditional χ^2-GOF procedure in all tested scenarios.

5.6 Evaluation of $\phi(S)$ in 'Experimental' Conditions

The section presents the results of a comparison between $\phi(S)$ and $\phi(S)_{true}$, in which the ensembles where generated using 'experimental' trajectories.

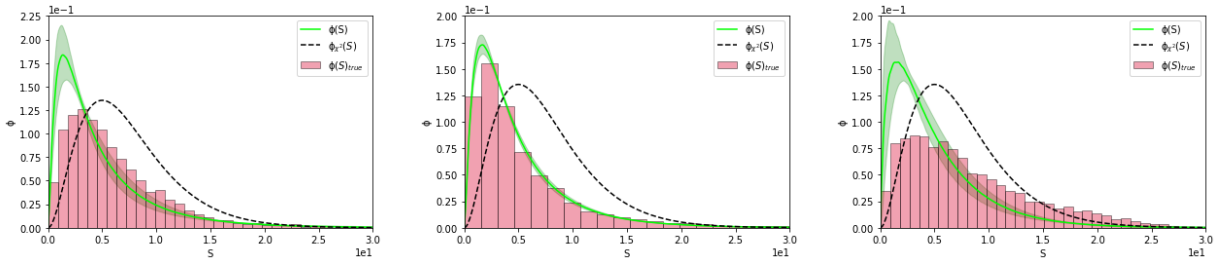
Chenouard et al in their paper 'Objective comparison of particle tracking methods' provide supplementary videos of noisifed and pixelated particle motion simulations [43], representing vesicle movement (BM), used for testing a given particle tracking method. The trajectories have been extracted from each supplementary video using ImageJ [44], and the '2D/3D Particle Tracker' plug-in developed by the MOSAIC group [12, 43].

The $\phi(S)$ evaluated using the 'experimental' trajectories is compared to the $\phi(S)_{true}$ generated via the BM prototype model for a given system. The results are presented in Figure 11. Further detail regarding the specifications and parameters of the '2D/3D Particle Tracker' plug-in can be found in the Supplementary Tables section of the Appendix (Table A).

Supplementary Video Specifications			
Supplementary Video	Vesicle Density	Trajectories	Sampling Points
S1	Medium	M = 27	N = 6
S5	Low	M = 283	N = 6
S6	High	M = 7	N = 6

Table 1: **Supplementary Video Specifications**. The amount of M captured trajectories for each supplementary video is detailed. In all cases, all trajectories with a minimum $N = 6$ sampling points where captured and used in ensemble generation.

Table 1 displays the specifications and number of captured trajectories for each of the supplementary videos.



(a) **S1 -** $\phi(S)$ **vs.** $\phi(S)_{true}$ (b) **S5 -** $\phi(S)$ **vs.** $\phi(S)_{true}$ (c) **S6 -** $\phi(S)$ **vs.** $\phi(S)_{true}$

Figure 11: **Evaluation of** $\phi(S)$ **in 'Experimental' Conditions -** $\phi(S)$ **vs.** $\phi(S)_{true}$. Figure 11 shows comparisons between $\phi(S)$ and $\phi(S)_{true}$ for supplementary videos S1, S5 and S6. Figure 11 (a) shows the comparison between $\phi(S)$ and $\phi(S)_{true}$ in a medium vesicle density setting where $M = 27$, Figure 11 (b) shows the comparison between $\phi(S)$ and $\phi(S)_{true}$ in a low vesicle density setting where $M = 283$, and Figure 11 (c) shows the comparison between $\phi(S)$ and $\phi(S)_{true}$ in a high vesicle density setting where $M = 7$. In all cases the captured trajectories have N = 6 sampling points. Note, one can notice an decrease in match quality between $\phi(S)$ and $\phi(S)_{true}$ as M decreases.

The new GOF procedure can correctly evaluate a $\phi(S)$ using 'experimental' trajectories, given M is large (Figure 11). In the case where M is large (Figure 11 (b)), almost exact match between $\phi(S)$ and $\phi(S)_{true}$ can be seen. As M decreases (Figures 11 (b), (c)), particularly when $M \longrightarrow 1$, the match between $\phi(S)$ and $\phi(S)_{true}$ significantly decreases.

The increased difference between $\phi(S)$ and $\phi(S)_{true}$ as the number of M trajectories captured decreases suggests the breakdown of the new GOF procedure when M is not large enough for the multivariate CLT to be valid. This is to be expected, as when the multivariate CLT is not valid, $\varrho(\mathbf{\Lambda}, \theta)$ cannot be properly approximated by Equation 3.8. Additional comparisons between $\phi(S)$ and $\phi(S)_{true}$ as $M \longrightarrow 1$ can be found in the Supplementary Images section of the Appendix (Figure E).

For an explanation of why match quality decreases with M, one must recall the multivariate CLT outlined in Section 3.1. For any given system, decreasing M will eventually lead to the approximation of $\varrho(\mathbf{\Lambda}, \theta)$ presented in Equation 3.8 to become invalid, as the multivariate CLT is only valid as $M \longrightarrow \infty$. This suggests that there is likely a region of parameter space in which M is too low for the new GOF procedure to be considered reliable (Section 5.7) .

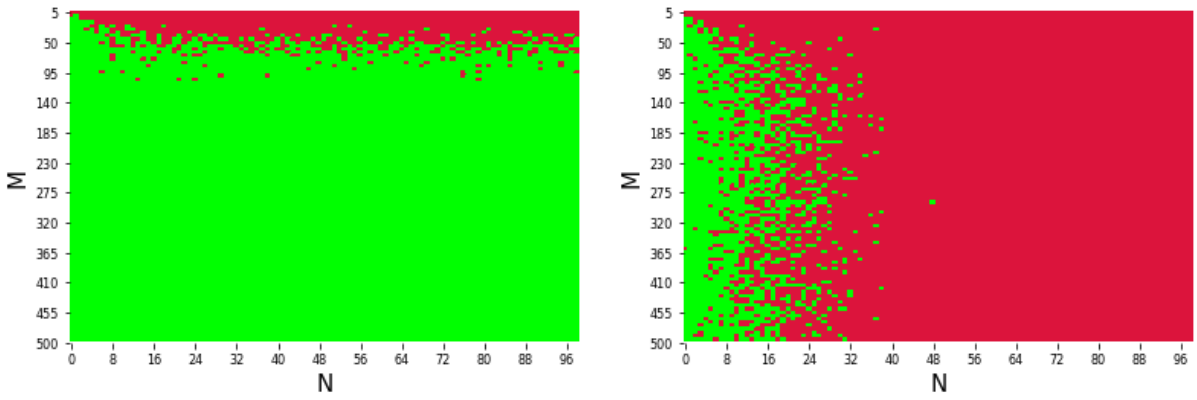
5.7 Reliable Parameter Space of the New GOF Procedure

From Section 5.6 it is apparent that the reliability of the new GOF procedure decreases with M. This section aims to demonstrate the region of reliable parameter space, in terms of M, for the new GOF procedure.

The combinations of M and N where the new GOF procedure is reliable can be determined by a phase-space plot. A phase-space plot, in this case, is an 100×100 matrix where each entry is given a colour dependant on its numerical value.

The Kolmogorov-Smirnov test statistic (Appendix D) is used to calculate a measure of the match (K) between the $F(S)$ and the cumulative distribution function of $\chi^2(\hat{F_{\chi^2}}(S))$ relative to the empirical $F(S)$ ($EMPF(S)$) generated from the $\phi(S)_{true}$ of a given system (Figure G), where $K \longrightarrow 0$ for a reliable match [45]. When $F(S)$ is a close match to $EMPF(S)$, a low K will be calculated, determining that, at a given M, the new GOF procedure is reliable.

The BM prototype model is used to evaluate $F(S)$, $\hat{F_{\chi^2}}(S)$ and the associated $EMPF(S)$ for range of systems covering the parameter space $M = 5 - 505$, in steps of five, and $N = 5 - 105$, in increments of one. K is then calculated for all systems throughout the selected parameter space for both $F(S)$ and $\hat{F_{\chi^2}}(S)$. The $M \times N$ matrix of K values is then transformed into a phase-space plot. The results can be found in Figure 12.



(a) $K(S)$ (b) $K_{\chi^2}(S)$

Figure 12: **Reliable Parameter Space of the New GOF Procedure - Phase-space Plots**. Figure 12 presents phase-space plots. Figure 12 (a) displays the reliable and unreliable regions of parameter space in terms of M and N for the new GOF procedure. Figure 12 (b) displays the reliable and unreliable regions of parameter space in terms of M and N for the traditional χ^2-GOF procedure. Green and red represent reliable and unreliable regions respectively. The cut-off value of K was set to $K = 0.3$, where if $K(S)$ or $K_{\chi^2}(S)$ has a $K > 0.3$, then the match is deemed unreliable. The raw formats of the phase-space plots presented can be found in the Supplementary Images section of the Appendix (Figure F). Typical fluctuations of a evaluated $\phi(S)$ and differences in $\phi(S)$ among function fitting methods can be found in Figure A and D respectively.

The phase space plot for the new GOF procedure shows a large region of reliable space, as the number of trajectories pass $M = 100$ (Figure 12 (a)). It is apparent that when $M \approx 100$, the multivariate CLT 'kicks in' and $\varrho(\Lambda, \theta)$ can be accurately approximated

by Equation 3.8. This can be seen by the large region of reliable parameter space after $M = 100$. Conversely, particularly so at high N, when trajectories decrease from $M = 100$ the regions of reliable parameter space rapidly disappear, with little to no reliable space as the number of trajectories drops below $M = 50$.

The reliable regions of parameter space for the χ^2-GOF procedure is much more narrow relative to that of the new GOF procedure (Figure 12 (b)). The reliability of the χ^2-GOF procedure decreases as N increases. There is a particularly sharp drop of as the number of sampling points passes $N = 10$. The small region of reliable space of $\phi_{\chi^2}(S)$ relative to that of $\phi(S)$ is again due to the difference in σ^2. As N increases, $\phi_{\chi^2}(S)$ becomes increasingly inaccurate as it deforms towards a traditional Gaussian with $\boldsymbol{\mu} = N$, with no ability to account for σ^2 variations caused by the correlations among the components of Λ. This in turn causes greater deviations between $\phi(S)$ and $\phi_{\chi^2}(S)$ as N increases.

Overall, it is apparent that $\phi(S)$ outperforms $\phi_{\chi^2}(S)$ in terms of reliability, with a larger region of reliable space, and a smaller K value in areas of overlapping reliability.

6 On the Application of $\phi(S)$

This thesis has primarily focused on the derivation and testing of $\phi(S)$. Here, how to use $\phi(S)$ in practice to determine the GOF of a given model is detailed.

Let us first define a scenario. Say one has a system of simulated or experimental trajectories, with which one can create an estimate ensemble average of some $\langle y(t) \rangle$, where the estimate ensemble average is made up of M trajectories of N sampling points. Say one has a model in mind with which to describe this estimate ensemble average, let us call this model A, and wishes to determine how consistent model A is with the trajectories that make-up the estimate ensemble average, i.e the GOF.

One could use the S (Equation 3.15) and $\phi(S)$ (Equation 3.26) defined in this thesis to determine the GOF of model A. The procedure is as follows:

- Firstly, one fits model A, and estimates $\boldsymbol{\Sigma}$ (Equation 3.6), $\mathbf{R}$ (Equation 3.13) and $\mathbf{P}$ (Equation 3.12).
- One can now extract the eigenvalues (λ) of $\mathbf{P}$ and evaluate the CF (Equation 3.24), this process is detailed in Section 3.5.
- At this point, one has enough information to evaluate $\phi(S)$ using the methods presented in sections 3.5 and 4.2.
- Now one has a suitable $\phi(S)$, the next step is to calculate S. Extracting the residuals around model A, i.e one must calculate $\boldsymbol{\Lambda}$ (Equation 3.7), allows for the calculation of S, as in Section 3.4.

- One now reviews where S falls on the associated $\phi(S)$, and makes a decision whether the model is valid or not.

In determining whether model A is a good fit, one would expect an S in the lower end of $\phi(S)$ in the case model A represents the data well, and an S in the tail end of the distribution in the case model A is a poor representation of the data.

In general, one would do the above steps for a small selection of possible models, say models A, B and C, using both S and $\phi(S)$ to determine the optimal model, solving the model selection problem one typically faces when fitting a model to an ensemble average.

7 Summary

In this thesis a new GOF procedure for testing the quality of a fitted model to time dependent ensemble averages has been presented. The new GOF procedure takes into account the correlations among the residuals ($\mathbf{\Lambda}$) of a fitted function around a ensemble average over time, a concept generally neglected. In this new GOF procedure, a test statistic (S) is calculated, where S is the normalised sum of square residuals (Equation 3.15), and its distribution ($\phi(S)$) is evaluated (Equation 3.26), where $\phi(S)$ takes into account the correlations among the components of $\mathbf{\Lambda}$. The new GOF procedure was tested under a variety of scenarios, $\mathbf{\Lambda}$ component correlation conditions and ensemble make-ups. In general, it was found that if M is large, that the new GOF procedure can be considered to be reliable, more so than the χ^2-GOF procedure in all tested scenarios. It was found that $\sigma^2(\phi(S)) > \sigma^2(\phi_{\chi^2}(S))$ for $N > 1$, which was found to be the driving factor in the increased reliability of the the new GOF procedure when compared to the χ^2-GOF procedure.

The tie between the reliability of the new GOF procedure and M is due to the multivariate CLT. When M is large, the multivariate CLT is valid, and the correlations among the components of $\mathbf{\Lambda}$ can be accurately approximated by the multivariate Gaussian presented in Equation 3.8. This is crucial, as when M is not large enough, the evaluated $\phi(S)$ is not an accurate estimation of $\phi(S)_{true}$. This can be seen in the phase-space plots (Figure 12), in which the new GOF procedure can be considered reliable after $M \approx 100$, past which the multivariate CLT has become valid.

When compared to the χ^2-GOF procedure, the new GOF procedure is a better estimate of $\phi(S)_{true}$ in all cases. This is down to the correlation among the components of Λ being non i.i.d (Figure 3). These non i.i.d correlations cause σ^2 to deviate from the theoretical prediction held by $\phi_{\chi^2}(S)$. This is apparent when one looks at the log-distribution of eigenvalues (r) of the Pearson correlation matrix ($\mathbf{P}$), denoted $\psi(r)_{true}$, where a much broader distribution of r than predicted by $\psi(r)$ can be observed (Figure 6). Where $\psi(r)$ is the typical distribution predicted by the Marchenko-Pastur law (Section 3.7), which is valid when the correlations among the components of $\mathbf{\Lambda}$ are indeed i.i.d. This broader distribution in turn suggests an increased σ^2. This increase in σ^2 causes the broadness

and asymmetry observed in $\phi(S)$, and is the key factor behind the deviation of $\phi(S)$ from $\phi_{\chi^2}(S)$, and its increased match quality when compared to $\phi(S)_{true}$. This difference can be observed in the phase-space plots (Figure 12), in which when $N > 10$ the difference in σ^2 is too great for $\phi_{\chi^2}(S)$ to be considered reliable.

Comparing briefly with current approaches in measuring GOF that consider correlations among the components of Λ (Section 3.4), it has been demonstrated that the new GOF procedure manages to avoid typical problems found in current approaches, producing suitable S regardless of dimensionality, and $\phi(S)$ provided $M > 100$.

This thesis fills the gap in traditional GOF testing procedures, arming scientists with a further approach to measure the GOF of a fitted model. The new GOF procedure considers the correlations among the components of $\mathbf{\Lambda}$, which are typically ignored, resulting in more reliable evaluations of the quality of a model fitted to an ensemble average. In communities where fitting models to ensemble averages is common practice, it is hoped the new GOF procedure will be used to aid in more accurate model selection.

References

[1] Brandenburg, B. and Zhuang, X., 2007. Virus trafficking – learning from single-virus tracking. Nature Reviews Microbiology, 5(3), pp.197-208.

[2] Liu, S., Wang, Z., Xie, H., Liu, A., Lamb, D. and Pang, D., 2020. Single-Virus Tracking: From Imaging Methodologies to Virological Applications. Chemical Reviews, 120(3), pp.1936-1979.

[3] Lakshmi, N. and Daniel, S., 2019. Physical Virology. Springer, Cham, pp.12-43.

[4] Saxton, M. and Jacobson, K., 1997. Single-particle Tracking:Applications to Membrane Dynamics. Annual Review of Biophysics and Biomolecular Structure, 26(1), pp.373-399.

[5] Notelaers, K., Smisdom, N., Rocha, S., Janssen, D., Meier, J., Rigo, J., Hofkens, J. and Ameloot, M., 2012. Ensemble and single particle fluorimetric techniques in concerted action to study the diffusion and aggregation of the glycine receptor 3 isoforms in the cell plasma membrane. Biochimica et Biophysica Acta (BBA) - Biomembranes, 1818(12), pp.3131-3140.

[6] Zelman-Femiak, M., 2022. Single Particle Tracking Membrane Receptor Dynamics. PhD. Julius-Maximilians-Universität Würzburg.

[7] Rizzo, M., Davidson, M. and Piston, D., 2009. Fluorescent Protein Tracking and Detection: Fluorescent Protein Structure and Color Variants. Cold Spring Harbor Protocols, 2009(12), p.pdb.top63.

[8] Foldes-Papp, Z. and Baumann, G., 2011. Fluorescence Molecule Counting for Single-Molecule Studies in Crowded Environment of Living Cells without and with Broken Ergodicity. Current Pharmaceutical Biotechnology, 12(5), pp.824-833.

[9] Taylor, R., Holler, C., Mahmoodabadi, R., Küppers, M., Dastjerdi, H., Zaburdaev, V., Schambony, A. and Sandoghdar, V., 2020. High-Precision Protein-Tracking With Interferometric Scattering Microscopy. Frontiers in Cell and Developmental Biology, 8.

[10] Ruthardt, N., Lamb, D. and Bräuchle, C., 2011. Single-particle Tracking as a Quantitative Microscopy-based Approach to Unravel Cell Entry Mechanisms of Viruses and Pharmaceutical Nanoparticles. Molecular Therapy, 19(7), pp.1199-1211.

[11] Tejedor, V., Bénichou, O., Voituriez, R., Jungmann, R., Simmel, F., Selhuber-Unkel, C., Oddershede, L. and Metzler, R., 2010. Quantitative Analysis of Single Particle Trajectories: Mean Maximal Excursion Method. Biophysical Journal, 98(7), pp.1364-1372.

[12] Sbalzarini, I. and Koumoutsakos, P., 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. Journal of Structural Biology, 151(2), pp.182-195.

[13] Fogelmark, K., Lomholt, M., Irbäck, A. and Ambjörnsson, T., 2018. Fitting a function to time-dependent ensemble averaged data. Scientific Reports, 8(1).

[14] Sivia, D. and Skilling, J., 2012. Data analysis. Oxford: Oxford University Press.

[15] Mecklin, C. and Mundfrom, D., 2007. An Appraisal and Bibliography of Tests for Multivariate Normality. International Statistical Review, 72(1), pp.123-138.

[16] Monnier, N., Guo, S., Mori, M., He, J., Lénárt, P. and Bathe, M., 2012. Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells. Biophysical Journal, 103(3), pp.616-626.

[17] Klauer, K., 2002. Model Testing and Selection, Theory of.

[18] Press, W., Teukolsky, S., Vetterling, W. and Flannery, B., 1992. Numerical recipes in C. Cambridge: Cambridge Univ. Pr.

[19] Werman, M. and Keren, D., 2001. A Bayesian method for fitting parametric and nonparametric models to noisy data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(5), pp.528-534.

[20] Rahman, G., Mubeen, S. and Rehman, A., 2022. Generalization of Chi-square Distribution. Journal of Statistics Applications Probability, 4, pp.119 - 126.

[21] Gil-Pelaez, J., 2022. Note on the inversion theorem.

[22] Imhof, J., 1961. Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika, 48(3/4), p.419.

[23] Davies, R., 1973. Numerical inversion of a characteristic function. Biometrika, 60(2), pp.415-417.

[24] Solomon, H. and Stephens, M., 1977. Distribution of a Sum of Weighted Chi-Square Variables. Journal of the American Statistical Association, 72(360), p.881.

[25] Liu, H., Tang, Y. and Zhang, H., 2009. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Computational Statistics Data Analysis, 53(4), pp.853-856.

[26] Witkovsky, V., 2016. Numerical inversion of a characteristic function: An alternative tool to form the probability distribution of output quantity in linear measurement models. ACTA IMEKO, 5(3), p.32.

[27] Duchesne, P. and Lafaye De Micheaux, P., 2010. Computing the distribution of quadratic forms: Further comparisons between the Liu–Tang–Zhang approximation and exact methods. Computational Statistics Data Analysis, 54(4), pp.858-862.

[28] Bodenham, D. and Adams, N., 2015. A comparison of efficient approximations for a weighted sum of chi-squared random variables. Statistics and Computing, 26(4), pp.917-928.

[29] Codling, E., Plank, M. and Benhamou, S., 2008. Random walk models in biology. Journal of The Royal Society Interface, 5(25), pp.813-834.

[30] Maly, I., 2002. Centrosome-dependant anisotropic random walk of cytoplasmic vesicles. Cell Biology International, 26(9), pp.791-799.

[31] Höfling, F. and Franosch, T., 2013. Anomalous transport in the crowded world of biological cells. Reports on Progress in Physics, 76(4), p.046602.

[32] Marquez-Lago, T., Leier, A. and Burrage, K., 2012. Anomalous diffusion and multifractional Brownian motion: simulating molecular crowding and physical obstacles in systems biology. IET Systems Biology, 6(4), pp.134-142.

[33] Goiko, M., de Bruyn, J. and Heit, B., 2018. Membrane Diffusion Occurs by Continuous-Time Random Walk Sustained by Vesicular Trafficking. Biophysical Journal, 114(12), pp.2887-2899.

[34] Fox, Z., Barkai, E. and Krapf, D., 2021. Aging power spectrum of membrane protein transport and other subordinated random walks. Nature Communications, 12(1).

[35] Johnson, R. and Wichern, D., 2007. Applied multivariate statistical analysis. 6th ed. New Jersey, USA: Pearson Prentice Hall.

[36] Mathai, A. and Provost, S., 1992. Quadratic forms in random variables. New York, NY: Dekker.

[37] Chaichian, M., 2019. Path integral in physics. [S.l.]: CRC Press.

[38] Pavez, E. and Ortega, A., 2021. Covariance Matrix Estimation With Non Uniform and Data Dependent Missing Observations. IEEE Transactions on Information Theory, 67(2), pp.1201-1215.

[39] Fan, J., Liao, Y. and Mincheva, M., 2011. High Dimensional Covariance Matrix Estimation in Approximate Factor Models. SSRN Electronic Journal.

[40] Hu, Z., Tong, T. and Genton, M., 2022. A Pairwise Hotelling Method for Testing High-Dimensional Mean Vectors. Statistica Sinica.

[41] T. Li, Y. and Wong, R., 2008. Integral and series representations of the dirac delta function. Communications on Pure amp; Applied Analysis, 7(2), pp.229-247.

[42] Yaskov, P., 2016. A short proof of the Marchenko–Pastur theorem. Comptes Rendus Mathematique, 354(3), pp.319-322.

[43] Chenouard, N., Smal, I., de Chaumont, F., Maška, M., Sbalzarini, I., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., Cohen, A., Godinez, W., Rohr, K., Kalaidzidis, Y., Liang, L., Duncan, J., Shen, H., Xu, Y., Magnusson, K., Jaldén, J., Blau, H., Paul-Gilloteaux, P., Roudot, P., Kervrann, C., Waharte, F., Tinevez, J., Shorte, S., Willemse, J., Celler, K., van Wezel, G., Dan, H., Tsai, Y., de Solórzano, C., Olivo-Marin, J. and Meijering, E., 2014. Objective comparison of particle tracking methods. Nature Methods, 11(3), pp.281-289.

[44] Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997-2018.

[45] Massey, F., 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association, 46(253), pp.68-78.

[46] Banna, O., Mishura, Y., Ralchencko, K. and Shklyar, S., 2019. Fractal Brownian Motion Approximations and Projections. Fractional Brownian Motion, pp.239-249.

Supplementary Tables

'MOSAIC' 2D/3D Particle Tracker Specifications			
Parameter	S1	S5	S6
Radius	2	2	2
Cutoff	0	0.006	0
Percentile	0.1	0.1	0.1
Link-range	1	1	1
Link-Length	10	10	10
Displacement	10	10	10
Dynamics	Brownian	Brownian	Brownian

Table A: **'MOSIAC' 2D/3D Particle Tracker Specifications**. The parameters used in the 'MOSIAC' 2D/3D particle tracker ImageJ plug-in when collecting trajectories in each of the supplementary videos are detailed. All supplementary videos depict noisified simulations of vesicle movement within membranes in a medium, low and high density setting respectively, with a signal to noise ratio (SNR) of 4 in all cases.

Supplementary Figures

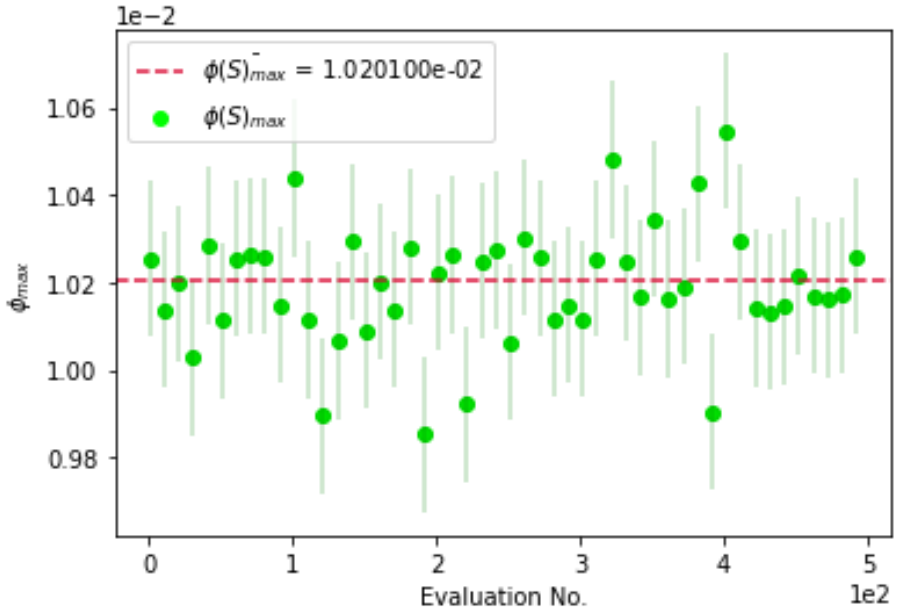
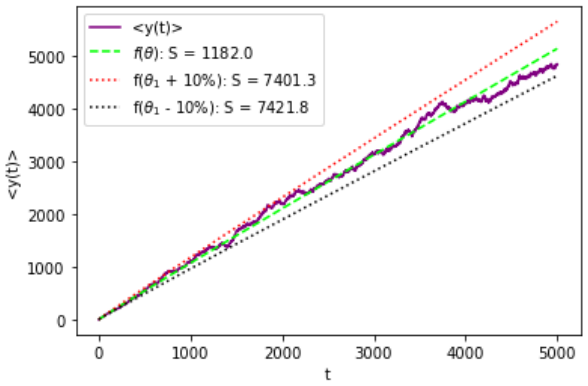
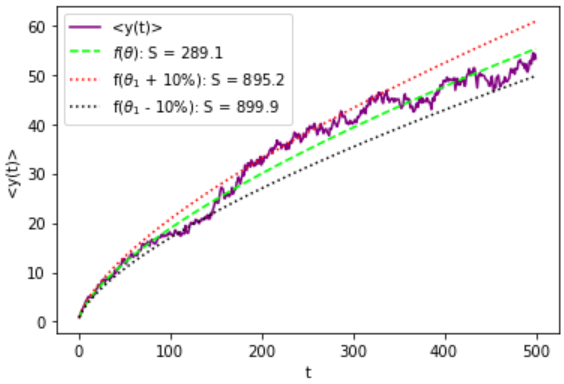


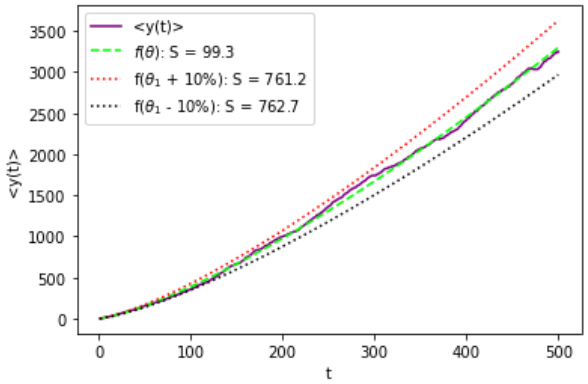
Figure A: **Typical Fluctuation Among $\phi(S)$ Evaluations**. Figure A presents the typical fluctuation in the evaluation of $\phi(S)$ for a given model. A series of $\phi(S)$ were evaluated using the BM prototype model with M = 1000 trajectories and N = 100 sampling points, $\phi(S)_{max}$ is then plotted against evaluation number. Here, $\phi(S)_{max}$ is taken as the maximum $\phi(S)$ value over all values of $\phi(S)$ for a given evaluation of $\phi(S)$.



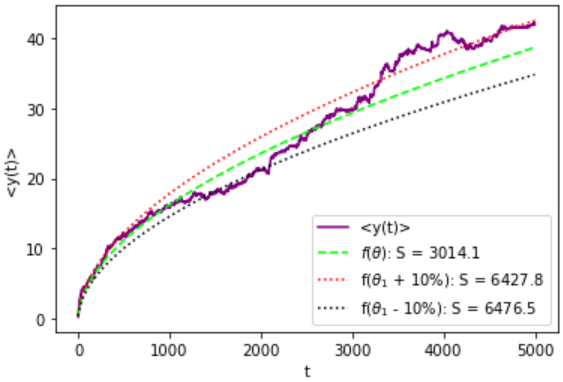
(a) **BM**



(b) **FRW (H = 0.33)**



(c) **FRW (H = 0.66)**



(d) **CTRW**

Figure B: **Typical Fits Among Prototype Models**. Figures B (a) - (d) present the typical fits associated with three different fit qualities for the BM (M = 250, N = 5000), FBM (H = 0.33, M = 250, N = 500), FBM (H = 0.66, M = 250, N = 500) and CTRW (M = 250, N = 5000) prototype models respectively. In case of the FBM and CTRW prototype models, the function fit is of the form, $f(\theta) = f(\theta_1, \theta_2) = \theta_1 t^{\theta_2}$, with $t = (1, ..., N)$. In the case of the BM prototype model, $f(\theta) = f(\theta_1) = \theta_1 t$. In all panels above, θ was estimated using the WLS-ICE method. Here, $t = (0, ..., N)$.

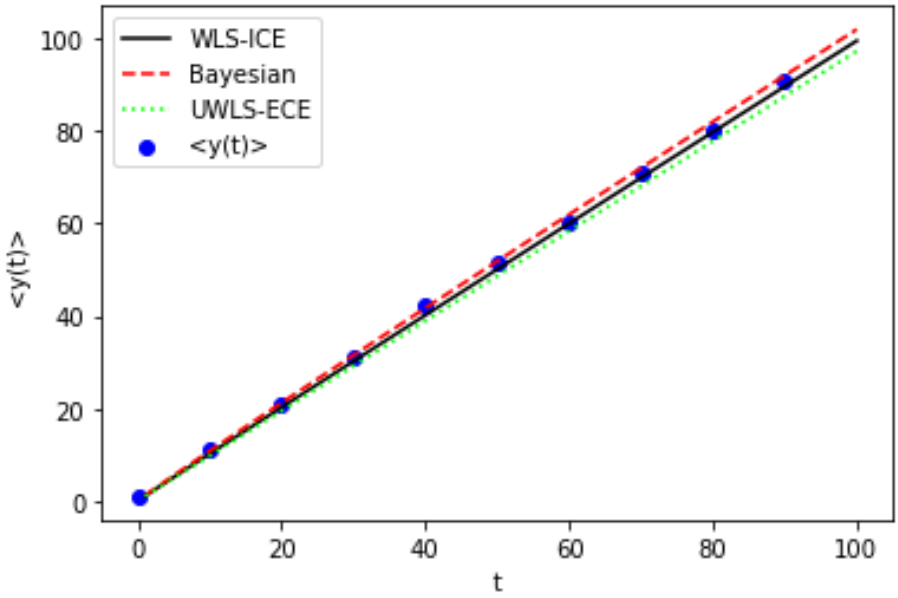
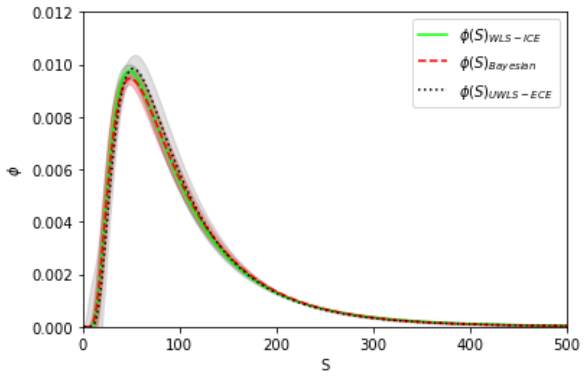
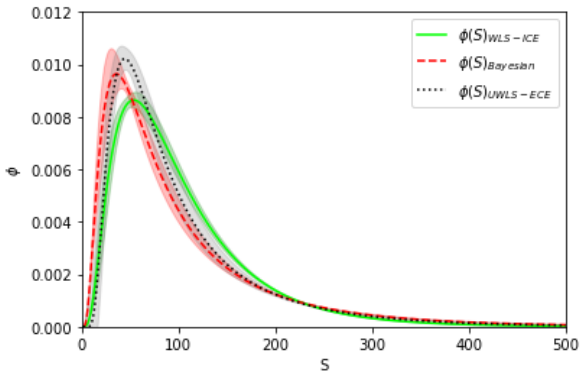


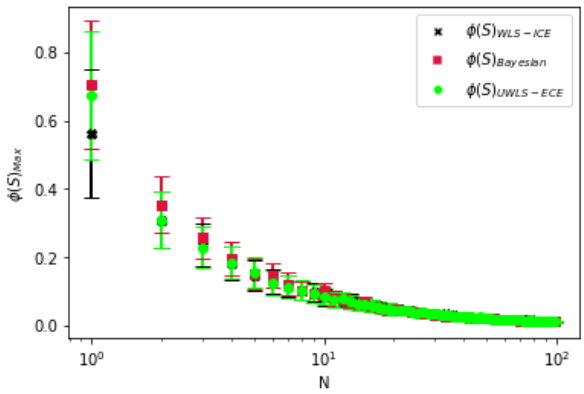
Figure C: **Comparison of Fitting Methods via Fitted Model**. Figure C displays the difference in fitted models among fitting methods. The test ensemble was generated using the BM prototype model, with M = 1000 trajectories and N = 100 sampling points, with $t = (0, ...N)$.



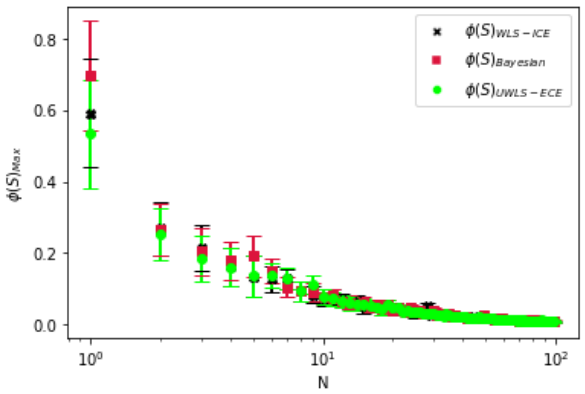
(a) **Comparison of $\phi(S)$ relative to choice of fitting method. M = 100.**



(b) **Comparison of $\phi(S)$ relative to choice of fitting method. M = 10.**

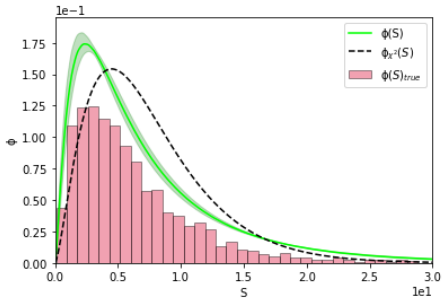


(c) **Comparison of $\phi(S)$ as a function of N, relative to choice of fitting method. M = 100.**

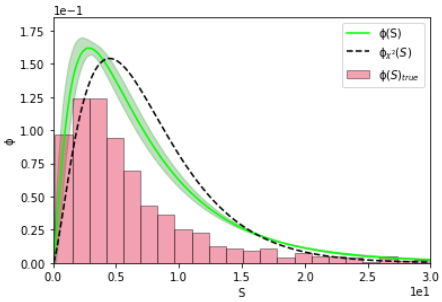


(d) **Comparison of $\phi(S)$ as a function of N, relative to choice of fitting method. M = 10**

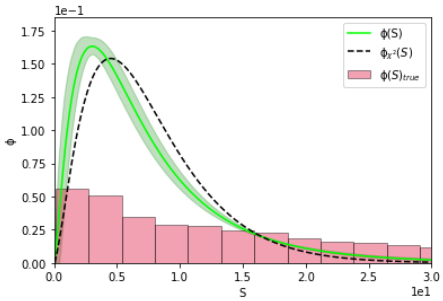
Figure D: **Fitting Method Bias Testing - $\phi(S)$ Evaluated Against Choice of Fitting Method**. Figure D presents $\phi(S)$ evaluated against fitting method under various scenarios. Figure D (a) presents fully evaluated $\phi(S)$ with a model fit using three different model fitting methods, WLS-ICE, Bayesian and UWLS-ECE, with the test ensemble generated from the BM prototype model with M = 100 trajectories and N = 100 sampling points. Figure D (b) presents a similar comparison, with the test ensemble in this case having M = 10 trajectories and N = 100 trajectories. Figures D (c) and (d) present the maximum $\phi(S)$ as a function of N for each model fitting method. In Figure D (c) the test ensemble had M = 100 trajectories and N = 100 sampling points. In Figure D (d) the test ensemble had M = 10 trajectories and N = 100 sampling points.



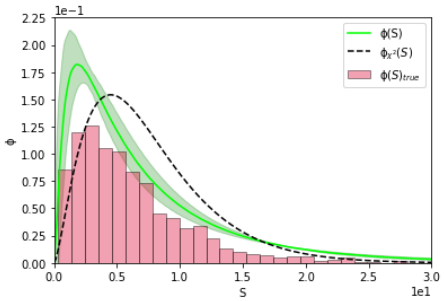
(a) **M = 96**



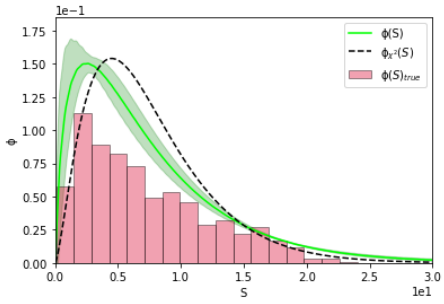
(b) **M = 48**



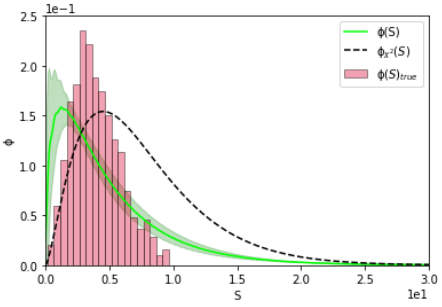
(c) **M = 24**



(d) **M = 12**

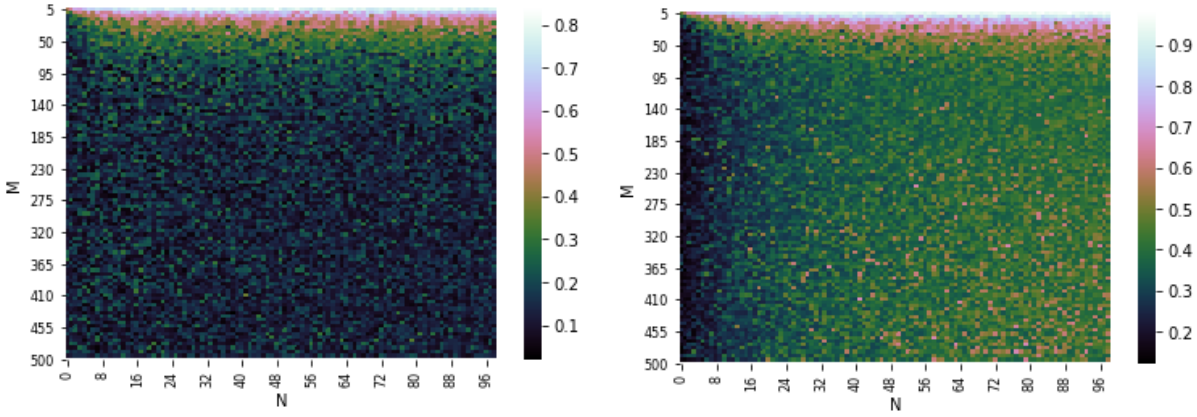


(e) **M = 6**



(f) **M = 3**

Figure E: **Supplementary Breakdown Analysis -** $\phi(S)$ **vs.** $\phi(S)_{true}$. Figure E displays supplementary test results comparing $\phi(S)$ and $\phi(S)_{true}$ at varied M. Figures E (a) - (f) compare $\phi(S)$ and $phi(S)_{true}$ when M = 96, M = 48, M = 24, M = 12, M = 6 and M = 3 respectively. In all cases the estimated ensemble was generated using the BM prototype model, with $N = 5$ sampling points.



(a) **K(S)** (b) **K(χ^2)**

Figure F: **Phase-space Plots - Raw Format**. Figure F presents the raw data format of the phase-space plots presented in Section 5.6 of the main text. Figure F (a) displays the numerical values of the Kolmogorov-Smirnov Test Statistic (K) for the specified combinations of M and N for the new GOF procedure. Figure F (b) displays the numerical values of K for the specified combinations of M and N for the χ^2-GOF procedure. In all cases the $EMPF(S)$ was generated with $z = 500$ calculated S. See Appendix D for further information regarding the calculation of K. Figure G provides an example of a typical K measurement.

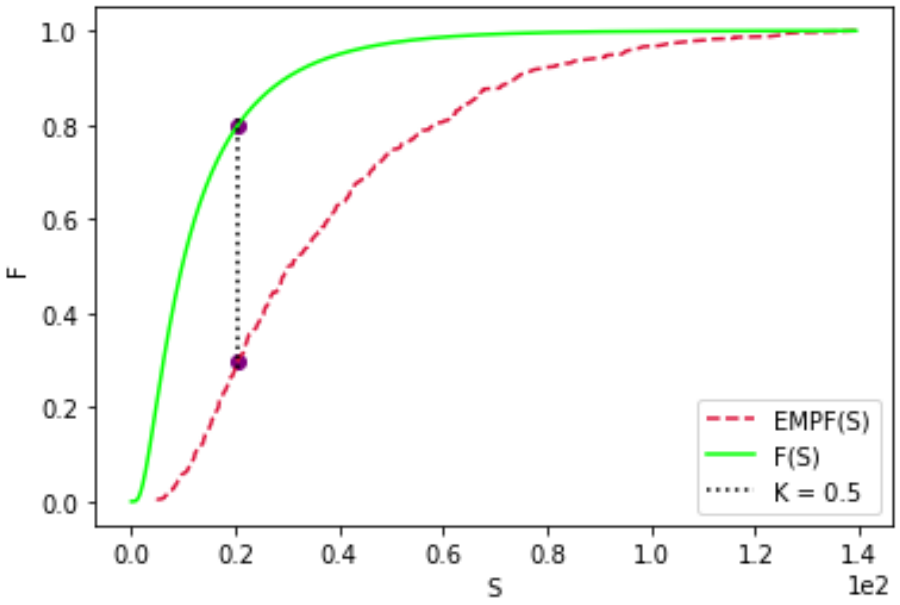
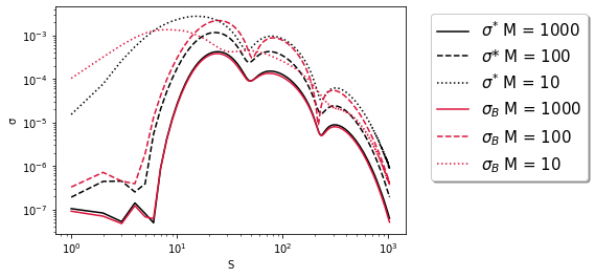
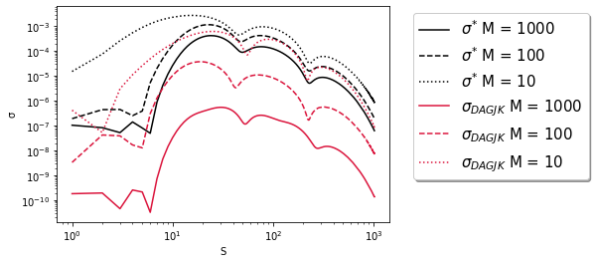


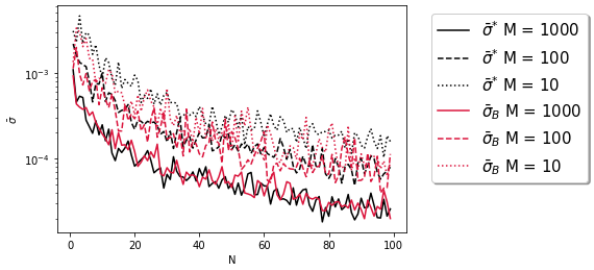
Figure G: **Typical K Measurement**. Figure G displays a typical measurement of the Kolmogorov-Smirnov test statistic (K) for a given system via comparison between $F(S)$ and $EMPF(S)$. The estimate ensemble was generated using the BM prototype model, with $M = 50$ trajectories and $N = 20$ sampling points. The EMPF(S) was generated through $z = 500$ calculated S.



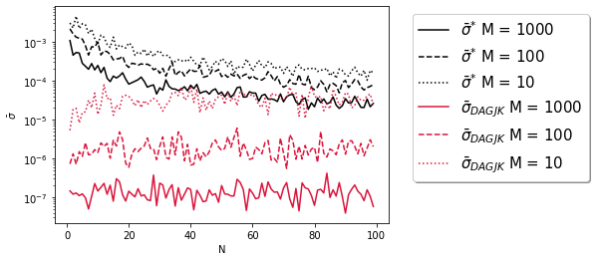
(a) σ_B **vs.** σ^*



(b) σ_{DAGJK} **vs.** σ^*

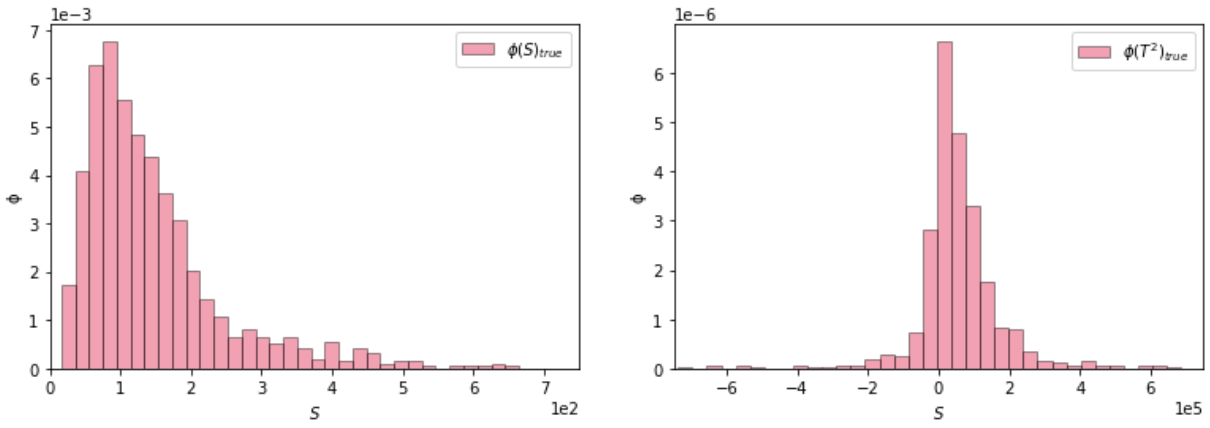


(c) $\bar{\sigma_B}$ **vs.** $\bar{\sigma^*}$ **as a Function of N**



(d) $\sigma_{D\bar{A}GJK}$ **vs.** $\bar{\sigma^*}$ **as a Function of N.**

Figure H: **Error Estimation Methods Comparison - σ_B vs. σ_{DAGJK} vs. σ^*.** Figure H presents the comparison between both σ_B and σ_{DAGJK} to σ^* at three different trajectory counts, $M = 10, 100$ and 1000. Figure H (a) displays the comparison of σ_B to σ^*. Figure H (b) displays the comparison of σ_{DAGJK} to σ^*. Figure H (c) presents $\bar{\sigma_B}$ vs $\bar{\sigma^*}$ as a function of N. Figure H (d) presents $\sigma_{D\bar{A}GJK}$ vs $\bar{\sigma^*}$ as a function of N. In all cases the estimate ensembles where generated with the BM prototype model. Figures H (c) and (d) have their x-axis in the log scale to aid with clarity. See Appendix F for detail on the estimation of σ_B and σ_{DAGJK}.



(a) $\phi(S)_{true}$ (b) $\phi(T^2)_{true}$

Figure I: $\phi(S)_{true}$ **vs.** $\phi(T^2)_{true}$. Figures I (a) and (b) present a typical evaluation of $\phi(S)_{true}$ and $\phi(T^2)_{true}$ respectively. In both cases the estimate ensembles were generated using the BM prototype model with $M = 1000$ and $N = 100$ sampling points. For a definition of $\phi(S)_{true}$, see the preamble of Section 5.

A Algorithm Derivations

In this section of the appendix, derivations of Equations 4.44 and 4.49 in the main text are presented.

A.1 Evaluating $\phi(S)$ - Derivation of Equation 4.44

Evaluation of $\phi(S)$ is achieved by applying the trapezoidal rule to the integral shown in Equation 4.43 [18, 20]. Gil-Pelaez provides,

$$\phi(S) = \frac{1}{\pi}\int_0^\infty R\{e^{(-ikS)}CF(k)\}dk. \tag{A.1}$$

where k is a real-valued Fourier variable in the range $[-\infty, \infty]$. Here, $k = (1, ..., n)$, where n is the selected number of points over which the integral is evaluated. Equation A.1 can be evaluated through use of the trapezoidal rule.

Let us now define a maximum and minimum value to the space in which we evaluate $\phi(S)$, noted as U and L respectively, where $S \in [L, U]$. One now has an area over which to evaluate, denoted by B, where $B = (L, ...U)$ and is of length n. The six-sigma rule can be used to select U and L, such that U - L spans a range where all possible S for a given system could fall [26], as such,

$$U = \sum_{i=0}^{N}\lambda_i + \sum_{i=0}^{N}\lambda_i^2, \tag{A.2}$$

and $L = 0$ by definition as $S > 0$ for all S.

Representing dk by, $dk = 2\pi/B$, one can evaluate Equation A.1 as follows,

$$\phi(S) = \frac{dk}{\pi}\left(\frac{1}{2} + \sum_{j=1}^{n} w_j \mathcal{R}\left(e^{-ik_jS}CF(k_j)\right)\right), \tag{A.3}$$

where w_j are the quadrature weights. In Equation A.3,

$$w_j = \begin{cases} \frac{1}{2} & j = 0, j = N \\ 1 & otherwise \end{cases}. \tag{A.4}$$

The leading $\frac{1}{2}$ in Equation A.3 arises as,

$$e^{ikS} = cos(kS) + isin(kS), \tag{A.5}$$

where, when $k = 0$,

$$cos(0) + i\sin(0) = 1, \tag{A.6}$$

thus as per the trapezoidal rule,

$$\phi(S) = \frac{\cos(0) + i\sin(0)}{2} = \frac{1}{2}. \tag{A.7}$$

Romberg's method is then applied. See Press et al. for further details [18].

A.2 Evaluating $F(S)$ - Derivation of Equation 4.49

Imhof provides an adaptation of the Gil-Pelaez theorem fit for numerical inversion and evaluation of $F(S)$ [20, 21],

$$F(S) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty k^{-1}\mathcal{I}\{e^{(-ikS)}CF(k)\}dk. \tag{A.8}$$

Equation A.8 can the again be approximated using the trapezoidal rule (Appendix A.1),

$$F(S) = \frac{1}{2} - \frac{dk}{\pi}(\frac{N-k}{2} + \sum_{j=0}^{n} w_j \mathcal{I}\left(\frac{e^{ik_jS}CF(k_j}{k_j}\right), \tag{A.9}$$

The leading term in Equation A.9 is a consequence of the fact that

$$\lim_{k\longrightarrow 0}\frac{CF(k)}{k} =< S > -B, \tag{A.10}$$

where $< S > = N$. The proof of Equation A.10 is provided by Witkovsky [26].

In the case of Equation A.9 Romberg's method is again applied [18].

B The Prototype Models

The theoretical background and implementation strategies of the three chosen prototype models are detailed in this section of the Appendix.

B.1 Brownian Motion

B.1.1 Brownian Motion Model Theory

BM is represented by a simple random walk model. The random walk is a zero μ process, with step sizes drawn from a Gaussian distribution, with the MSD being described by

$$y(t_i) =\mid y^{(m)}(t_i) - y^{(m)}(0) \mid^2 . \tag{B.11}$$

Analytically the covariance among squared displacements is given by,

$$Q^*_{ij} = \langle (y(t_i) - \langle y(t_i) \rangle) (y(t_j) - \langle y(t_j) \rangle) \rangle = 2 (C_{ij})^2 = 2 \times (2Dmin(t_i t_j))^2, \quad \text{(B.12)}$$

or can be estimated via Equation 3.11.

B.1.2 Brownian Motion Model Implementation

Each M trajectory is generated by a cumulative sum of jumps drawn from a Gaussian distribution, with zero μ and a chosen step length σ^2 (a^2). The number of jumps summed is determined by the chosen number of N sampling points, with the step increment (ϵ). This above is repeated M times and an ensemble generated according to Equation 4.51.

Unless otherwise stated, the BM prototype model has the following variables: $a^2 = 1$, $\mu = 0$, $\epsilon = 1$. The choices of M and N vary between simulations and are noted in text.

B.2 Fractal Brownian Motion

B.2.1 Fractal Brownian Motion Model Theory

FBM is a zero μ Gaussian process, in which the increments between steps are not independent. Each step can be related via the auto-correlation function,

$$C^*_{ij} = \frac{1}{2}(t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H}), \quad \text{(B.13)}$$

where H is the Hurst parameter defining the type of dependence, that being linearly ($H = 1/2$), positively ($H > 1/2$) or negatively correlated ($H < 1/2$) relative to time, with $H = 1/2$ describing BM [13, 42]. Where, when $H = \frac{1}{2}$,

$$(t_i^{2\frac{1}{2}} + t_j^{2\frac{1}{2}} - |t_i - t_j|^{2\frac{1}{2}}) = \frac{1}{2}(t_i + t_j - |t_i - t_j|) = 2\frac{1}{2}min(t_i t_j) = 2Dmin(t_i t_j), \quad \text{(B.14)}$$

which describes the covariance of trajectories in BM, given D = $\frac{1}{2}$.

B.2.2 Fractal Brownian Motion Implementation

The FBM prototype model follows the implementation method detailed by Hosking [46]. Each M trajectory is again a cumulative sum of jumps drawn from a Gaussian distribution, though in the case of FBM, μ and a^2 are updated at each increment using the auto-correlation function defined in B.14. This update procedure is completed N times for M given trajectories. The ensemble average is then given by Equation 4.51.

All variables within the FBM prototype model vary between simulation and are therefore noted in text relative to each simulation detailed. This is with the exception of ϵ, where $\epsilon = 1$ throughout.

B.3 Continuous Time Random Walk

B.3.1 Continuous Time Random Walk Theory

CTRW is a zero μ Gaussian process, with jump sizes drawn from a Gaussian distribution with a chosen a^2. Each M trajectory is a cumulative sum of N jumps, with random a given wait time between each jump (t_{lag}), with t_{lag} generated through,

$$t_{lag} = r^{\frac{-1}{\alpha}} - 1. \tag{B.15}$$

B.3.2 Continuous Time Random Walk Implementation

The implementation procedure for the CTRW prototype model is similar to that of the BM prototype model, where a given trajectory is created via a cumulative sum of jumps drawn from a Gaussian distribution with a chosen a^2, though, in this case the time increment between in jump is derived from Equation B.15. As such, if $t_{lag} > \epsilon$, then subsequent jumps are are taken to be zeros, and the current position of the trajectory remains the same until the next jump takes place following the expiry of the lag time, or until the chosen number of N is exceeded.

The variable settings within the CTRW prototype model are as follows: $\alpha = 0.5$, $\epsilon = 1$ and $a^2 = 1$. The number of N sampling points, as well as the total number M trajectories, varies between simulations, as such each simulation detailed in text has the associated N and M noted.

C Marchenko-Pastur Law

The Marchenko-Pastur (MP) law gives a typical distribution of eigenvalues ($\psi(\lambda)$) for a random matrix, where for the MP law to be valid, the random matrix must be made up of i.i.d sample vectors (Equation 3.38). $\psi(\lambda)$ is given by,

$$\psi(\lambda) = \begin{cases} \dfrac{1}{2\pi\lambda\sqrt{\frac{N}{M}}}\sqrt{(b-\lambda)(\lambda-a)} & a < \lambda < b \\ 0 & otherwise \end{cases}, \tag{C.16}$$

where a and b are defined as follows,

$$a = (1 - \sqrt{\frac{N}{M}})^2, \tag{C.17}$$

$$b = (1 + \sqrt{\frac{N}{M}})^2, \tag{C.18}$$

where a and b also define the lower and upper bounds of $\psi(\lambda)$ respectively.

In the main text (Figure 6), $\psi(r)$, where $r = \ln(\lambda)$ is used in place for $\psi(\lambda)$, $\psi(r)$ is derived as follows,

$$\psi(r) = \int_a^b \delta(r - r')\psi(\lambda')d\lambda'. \tag{C.19}$$

At this point, a change of coordinates is introduced,

$$\lambda' = e^{r'}d\lambda' = e^{r'}dr'. \tag{C.20}$$

Continuing,

$$\psi(r) = \int_{\ln(a)}^{\ln(b)} \delta(r - r')\psi(e^{r'})e^{r'}dr' = \begin{cases} \psi(e^r)e^r & \ln(a) < \lambda < \ln(b) \\ 0 & otherwise \end{cases}, \tag{C.21}$$

evaluating $\psi(e^r)e^r$ leads to,

$$\psi(r) = \begin{cases} \dfrac{1}{2\pi\sqrt{\frac{N}{M}}}\sqrt{(b - e^r)(e^r - a)} & \ln(a) < \lambda < \ln(b) \\ 0 & otherwise \end{cases}. \tag{C.22}$$

In the case where $M \longrightarrow \infty$,

$$a = (1 - \sqrt{\frac{N}{M}})^2 \longrightarrow 1, \tag{C.23}$$

$$b = (1 + \sqrt{\frac{N}{M}})^2 \longrightarrow 1. \tag{C.24}$$

In this case one would expect $\lambda = 1$ for all λ, leading to

$$CF(k) = (1 - 2ik)^{\frac{N}{2}}, \tag{C.25}$$

which is the CF associated with $\phi_{\chi^2}(S)$ (Equation 3.25).

D Kolmogorov-Smirnov Test

This section of the Appendix describes the Kolmogorov-Smirnov (KS) test statistic (K) and it's use in generating the phase-space plots of the reliability of the new GOF procedure in parameter (M, N) space (Figure 13).

K is a measure of the statistical distance between two ϕ that span the same probability space [45]. K can thus be used as measure of how well $\phi(S)$ matches $\phi(S)_{true}$ in a given setting,

$$K = \max_{1 \leq i \leq n} \{F(S)_i - EMPF(S)_i, EMPF(S)_i - F(S)_i\}, \tag{D.26}$$

where $EMPF(S)$ denotes the empirical $F(S)$.

Generating a series of K for a variety of M and N combinations allows the generation of a phase space plots of K values in (M, N) coordinates, allowing for a measure of reliability in different scenarios, where $K \longrightarrow 0$ as reliability increases.

In all cases $\phi(S)_{true}$ was generated with $z = 500$ S values using the BM prototype model. The phase-space plots presented in text are 100×100 matrices of K values associated with a selected M and N. Regions of low K represent regions of high reliability, and visa-versa.

E Function Fitting Methods

This section of the Appendix details the fitting methods referenced throughout the thesis.

E.1 Weighted-Least-Squares Including Correlation Error

The Weighted-Least-Squares Including Correlation Error (WLS-ICE) method is the primary fitting method used throughout the thesis.

In all cases presented a model based on a power law is fit, as such the WLS-ICE estimates two parameters (θ_1, θ_2) in order to fit a time dependent power law,

$$f(\theta, t)_{WLS-ICE} = \theta_1 t^{\theta_2}, \tag{E.27}$$

where $t = (1, ..., N)$. The optimal parameters are achieved through minimisation of S defined in Equation 3.16. Minimisation of S is achieved by solving $\frac{\delta S}{\theta_a} = 0$,

$$\frac{\delta S}{\theta_a} = 2 \sum_{i,j} \frac{\delta f_i(\theta)}{\delta \theta_a} R_{i,j} (f_i(\theta) - \bar{y_j}) = 0, \tag{E.28}$$

where R is as defined in Equation 3.15 and a spans J free fitting parameters, $a = (1, ..., J)$.

E.2 Bayesian Regression

Bayesian regression aims to maximise the likelihood function of a chosen parameter, in this case fitting parameters.

Estimation of optimal fitting parameters can be achieved by maximising the probability, via Bayes formula,

$$P(f(\theta, t)|\hat{y}) = \frac{P(\hat{y}|f(\theta, t))P(f(\theta, t))}{P(\hat{y})} \propto \prod_{i=1}^{N} P(y_i^{(m)}|f(\theta, t)) \tag{E.29}$$

where $\hat{y}$ defines the ensemble average, made up of $y^{(m)} = (y_1^{(m)}, ..., y_N^{(m)})$ points along a given m trajectory, where $m = (1, ..., M)$. One can represent $P(y_i^{(m)}|f(\theta,t))$ by the following integral,

$$P(y_i^{(m)}|f(\theta,t)) = \int_{f(\theta,t)} P(y_i^{(m)}|p)(p|f(\theta,t))dp \tag{E.30}$$

where p is a point on the fitted model. The model that maximises Equation E.30 is the optimal fit for a given ensemble.

F Error Estimation Methods

The section of the Appendix details the two methods used to estimate the standard deviation, σ, of an evaluated $\phi(S)$, this serves as the error on all $\phi(S)$ plots in the main text.

F.1 Bootstrap Error Estimation

Bootstrap (B) sampling with replacement can be used to estimate the σ of a given parameter.

B sampling creates 'new' trajectories via re-sampling already existing data. In a given system, the M trajectories are sampled at random with replacement, meaning a given trajectory can be selected more than once. This is done M times creating a 'new' ensemble. This 'new' ensemble can then be used to evaluate a 'new' $\phi(S)$, noted $\phi(S)_B$.

Repetition of B sampling b = 100 times allows for b $\phi(S)_B$ to be evaluated, leading to the following estimation of σ for a given point on $\phi(S)$.

$$\sigma_B = \sqrt{\frac{\sum_{i=0}^{b}(\phi(S)_i - \bar{\phi(S)})^2}{b}}. \tag{F.31}$$

Equation F.31 can be used to estimate σ at every point along a given $\phi(S)$. Unless otherwise stated, σ_B was estimated using b = 100 evaluations of $\phi(S)_B$. σ_B then provides an suitable estimate of σ^* for a given system evaluated $\phi(S)$.

A comparison between σ_B and σ^* can be found in Figure H.

F.2 Delete-a-Group-Jackknife Error Estimation

Delete-a-group-jackknife (DAGJK) sampling can be used to estimate σ of a given point of $\phi(S)$.

DAGJK sampling refers to the principle of splitting up a given data set into groups, systematically repeating a given procedure sequentially removing a given group from the data set on each repetition of the procedure. DAGJK sampling provides an estimate of σ by fist splitting the trajectories making up a given ensemble into g groups of h trajectories (where $M = g \times h$).

DAGJK sampling allows for g new estimates of $\phi(S)$, $\phi(S)_{DAGJK}$, leading to the following estimator of σ for a given point on $\phi(S)$

$$\sigma_{DAGJK} = \sqrt{\frac{g-1}{g}\sum_{i=0}^{g}(\phi(S)_i - \bar{\phi(S)})^2}. \tag{F.32}$$

Equation F.32 can be used to estimate σ at every point along a given $\phi(S)$. Unless otherwise stated, σ_{DAGJK} was evaluated with $g = 10$ and $h = M/10$.

A comparison between σ_{DAGJK} and σ^* can be found in Figure H.