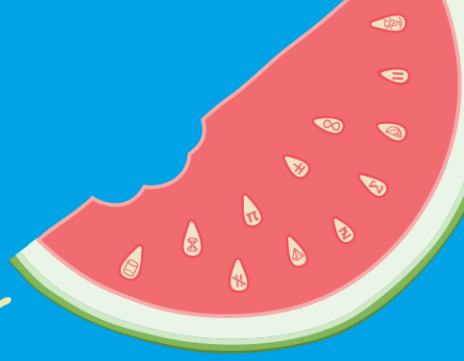
AMSI VACATION RESEARCH SCHOLARSHIPS 2021–22

Get a taste for Research this Summer



Extreme Events and Critical Fluctuations in Generated Time Series Data

Daniel Claassen

Supervised by Thomas Stemler

University of Western Australia

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Abstract

Methods available to predict phase transitions have not been tested rigorously in a simulated environment, and fail to accurately predict them in the real world. Our goal was to discover whether generated data exhibits the kinds of indicators we have observe in real-world data, and whether this kind of data holds up to typical methods of analysis. We generated data from the symmetric double-well potential system with noise forcing utilising the Euler-Maruyama method. Then we attempted to characterise the 2 phase transitions with auto correlative and variance analyses. We were successful in identifying the second phase transition, a supercritical pitchfork bifurcation. As an extension of this general question and methodology, we extended our analysis to data with a lower time resolution, as well as real world ice core data, and found similar indicators of phase transition.

Contents

1 Introduction **3**

2 Statement of Authorship **3**

3 Systems of Interest **3**

3.1 Preface . . . 3

3.2 The Symmetric Double-Well Potential . . . 4

4 Generation of Time Series Data **4**

4.1 Euler Maruyama Method . . . 4

4.2 Implementation . . . 6

5 Results and Discussion **6**

5.1 Methods of Analysis . . . 6

5.2 Expectations . . . 7

5.3 Analysis . . . 7

5.4 Hysterisis . . . 9

6 Further Applications **10**

6.1 Lowered Resolution . . . 10

6.2 Ice Core Data . . . 12

7 Conclusions **13**

References **15**

8 Appendices **15**

8.1 Appendix 1: . . . 15

8.2 Appendix 2: . . . 16

8.3 Appendix 3: . . . 17

8.4 Appendix 4: . . . 17

8.5 Appendix 5: . . . 18

8.6 Appendix 6: . . . 18

1 Introduction

Long form time-series data acquired across many fields enables us to identify where regime changes and phase transitions occur. Understanding what causes or enables these transitions to occur, and how they are established, is of extreme importance to fields such as climatology- where climate change is enabling many of the systemic behaviours traditional evidence would say leads to extreme phase transitions. Phase transitions are one of the best methods we have for understanding the behaviour of a complex system, and resultant dynamical behaviour.

Critical behaviour close to these transitions has certain characteristics (for example, extreme fluctuations) that we should be able to detect with a traditional analysis. If we were able to detect and characterise this extreme behaviour with absolute certainty, we would be able to predict what kind of transitions/bifurcations a given system is undergoing. As well as the severity and immediacy of such transitions.

Early warning signs are typically thought of as the manifestation of this extreme behaviour, though they can appear in different ways. Physically in a real-world system, abnormal, strenuous, or significantly odd behaviour can be thought of as an early warning sign. For example, an increasing commonality of extreme weather events. Though they themselves are extreme behaviour, the periodicity with which they occur may signal a loss in robustness or the recovery rate of climate system, which in turn is an early warning sign of a phase transition.

2 Statement of Authorship

The work was divided as follows,

- Daniel Claassen was responsible for developing the Julia code that produced the results, reported and interpreted results, and wrote this report.
- Thomas Stemler was responsible for providing direction, avenues of research and relevant literature, checking code, supervising the structure of work, and proofreading this report.

3 Systems of Interest

3.1 Preface

Complex Systems that are capable of having their behaviours described analytically are often difficult to simulate due to the interaction of many variables in continuous time. They display a rich variety of dynamical behaviour and this makes it difficult to detect phase transitions. Therefore, it is important that we strike a balance- when we consider which systems to investigate, between the costs of simulation and analysis against the gains of information and adherence to real-world behaviour.

For the purpose of identifying phase transitions, a given system must exhibit a robust, stable behaviour for a particular control parameter selection, as well as contain a large enough systematic change between phases

so as to be identifiable. Ideally, the system would also be able to be described with some form of Stochastic Differential Equation (SDE).

The system we will investigate is from the field of physics, the Symmetric Double-Well Potential system.

3.2 The Symmetric Double-Well Potential

Taken from the field of physics, the symmetric double-well potential system is a text book example given by:

$$F = -\nabla V \qquad\qquad V = cx^4 - bx^2 \tag{1}$$

F represents the force on a a particle, and V is the potential, and c and b are parameters. Taking the derivative on both sides we get $F = -\dot{V} = 2bx - 4cx^3$. This derivation describes the forces acting on the potential energy of a particle in a system.

This system has a simple and robust behaviour and is suitable for our purposes of analysis. We take $c = \frac{1}{4}$ and $b = \frac{-a}{2}$ in Equation 1, which allows variation in the system behaviour by changing a single control parameter, a.

Now consider the addition of a Brownian forcing to a particle moving along this field. Evident is the increasing potential barrier as a increases. Fluctuations could force the particle to move from one well to another.

In this scenario, this particular system has 3 phases displayed in Figure 2. The first phase, **I**, is the phase where the field is a uniform, symmetric monostable well. The behaviour would be governed entirely by noise and limited on both sides by the basin of the field. The second phase, **II**, is the phase where the particle is able to move over the potential barrier if the noise permits. The third phase, **III**, has two metastable points on either side of a large potential barrier that the particle is unable to cross. It occurs when the control parameter is too high for a fixed Brownian motion to overcome.

This system, alongside the addition of Brownian motion is the system we will generate data on.

4 Generation of Time Series Data

4.1 Euler Maruyama Method

The SDE to be integrated is:

$$\dot{x} = f(x) + \sqrt{D}\xi \qquad\qquad f(x) = ax - x^3 \tag{2}$$

The standard method for integrating such a SDE is the Euler-Maruyama method given by Herzel (1991).

$$x(t+\tau) = x(t) + \tau f(x) + dW \tag{3}$$

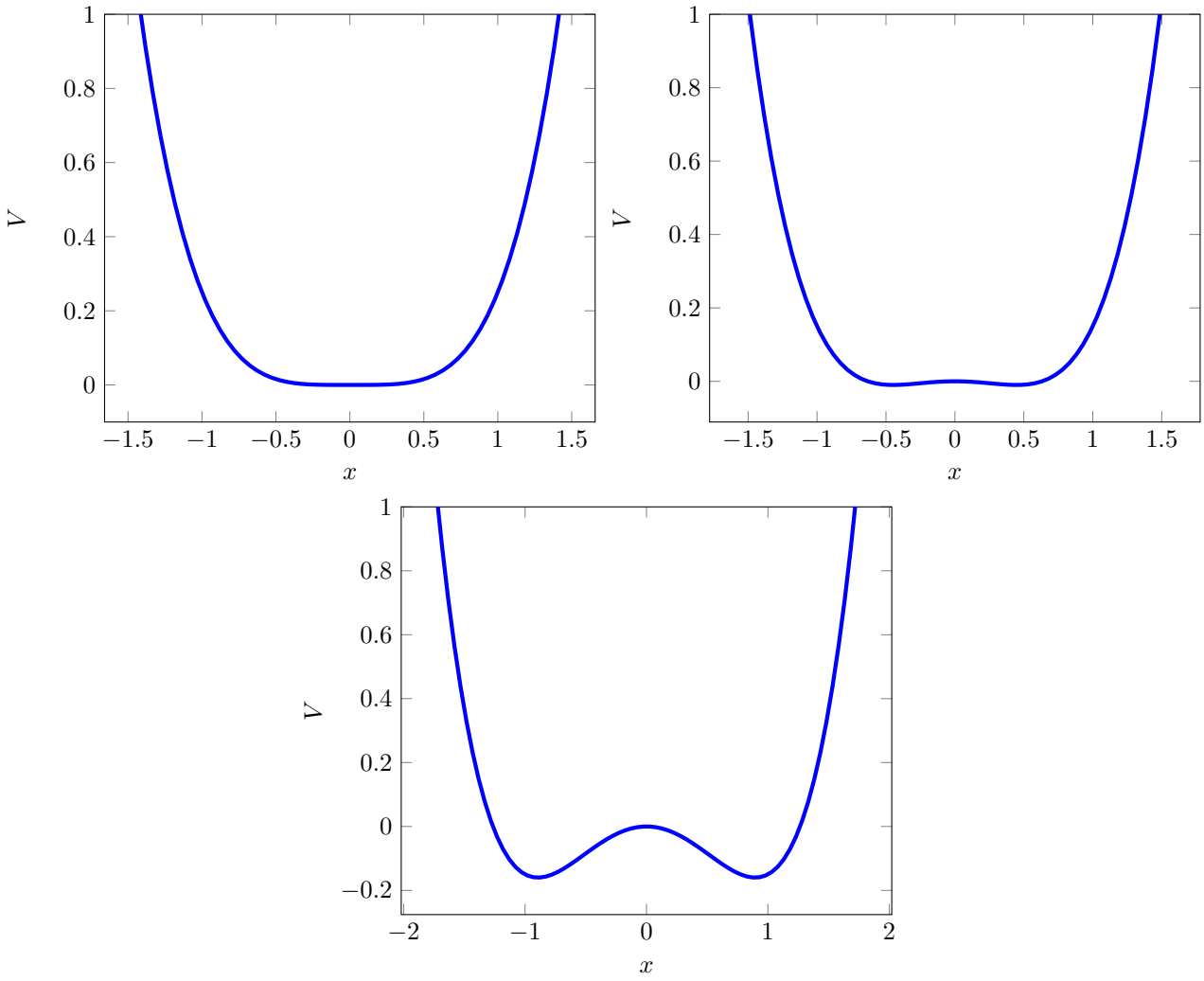


Figure 1: Cartesian plot of Eq.1 with $a = 0$ (monostable), $a = 0.4$ (bistable with small potential barrier), $a = 0.8$ (bistable with a high potential barrier)

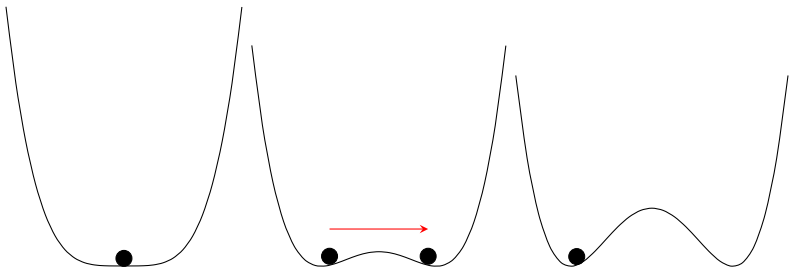


Figure 2: Higher control parameter values prevent the movement of a Brownian particle over the potential barrier. Left to right: Phase **I**, **II** and **III**

Where τ is a time increment dt, $t \longleftarrow t + \tau$ and dW is the Wiener Increment:

$$dW = \sqrt{2D\tau}(\xi - \xi') \tag{4}$$

Here, $2D$ is the variance of our Brownian Motion, and ξ/ξ' are two noise realisations drawn from the distribution $N(0, 1)$.

4.2 Implementation

The above was coded in Julia, and can be found in *Appendix 1.* We used a time increment $dt = \tau$ of 10^{-3}, an initial condition for x from the distribution $U(-1, 1)$, a D of 0.2, and a runtime for t of $[0, 10^4]$. Also of note is that the simulations in some cases exhibited a transient behaviour dependent upon initial condition. This behaviour was eliminated by removing early datapoints, which gives a new runtime of $[100, 10^4]$ These choices are largely arbitrary, and effect mostly the computational cost of the simulation. A smaller time increment or longer runtime would increase precision but effects are diminutive compared to costs. The transient behaviour cutoff, and noise was arbitrary.

The aim was to find the phases of the system empirically as well as their transitions. It is thus necessary to generate many of these simulations across a control parameter range in order to identify where measures in the system begin to change according to phase transitions. Generating data in large quantities was done using the code in *Appendix 2.* This writes our simulations to CSV files to be stored for analysis later. The range and density of control parameter simulations depended upon repeat testing and searching for phase transitions. For the particular set of parameters discussed above, we found that an interval of a along $[-0.1, 0.9]$ was suitable and covered all phases of the system. Increased ranges would not be useful as the phase transitions are all contained well within these bounds. The resolution would initially be low, so as to speed up the finding of phase transition behaviour, before the resolution would be increased to find smaller details.

5 Results and Discussion

5.1 Methods of Analysis

Many measures exist for analysing phase transitions in time series data. The primary measures are variance, and lag-1 auto-correlation (AC1). These have been used to great effect in the literature. Two examples of their use- variance, to predict the onset of seizures in the brain by Litt et al. (2001), and AC1 to determine climate transition dynamics by Dakos et al. (2008). Two other methods used in a similar fashion include the recovery rate of a given system, as well as the basin of attraction as established in Scheffer et al. (2012).

The methods we use in our analysis are variance and AC1. Recovery rate as a measure requires a more sophisticated analysis beyond the scope of this paper, and modelling the basin of attraction is unnecessary as the system has been defined in a closed system- we know the SDE governing it and can thus produce accurate representations of the basin of attraction.

5.2 Expectations

Scheffer et al. (2009) have shown that by modelling an SDE as an autoregressive equation, it is possible to show analytically that close to an established phase transition, variance is expected to diverge to infinity, and AC1 is expected to converge to 1. The derivation leading to this conclusion in their paper can be found in *Appendix 3*. We thus expect similar behaviours in our system.

5.3 Analysis

An example of the resultant output from the generation code at *Appendix 4* can be found in Figure 3. These are individual simulations run from the code at individual control parameter values. Each of these simulations had their variance and AC1 recorded respectively. These recordings would be taken together with other simulations at other values.

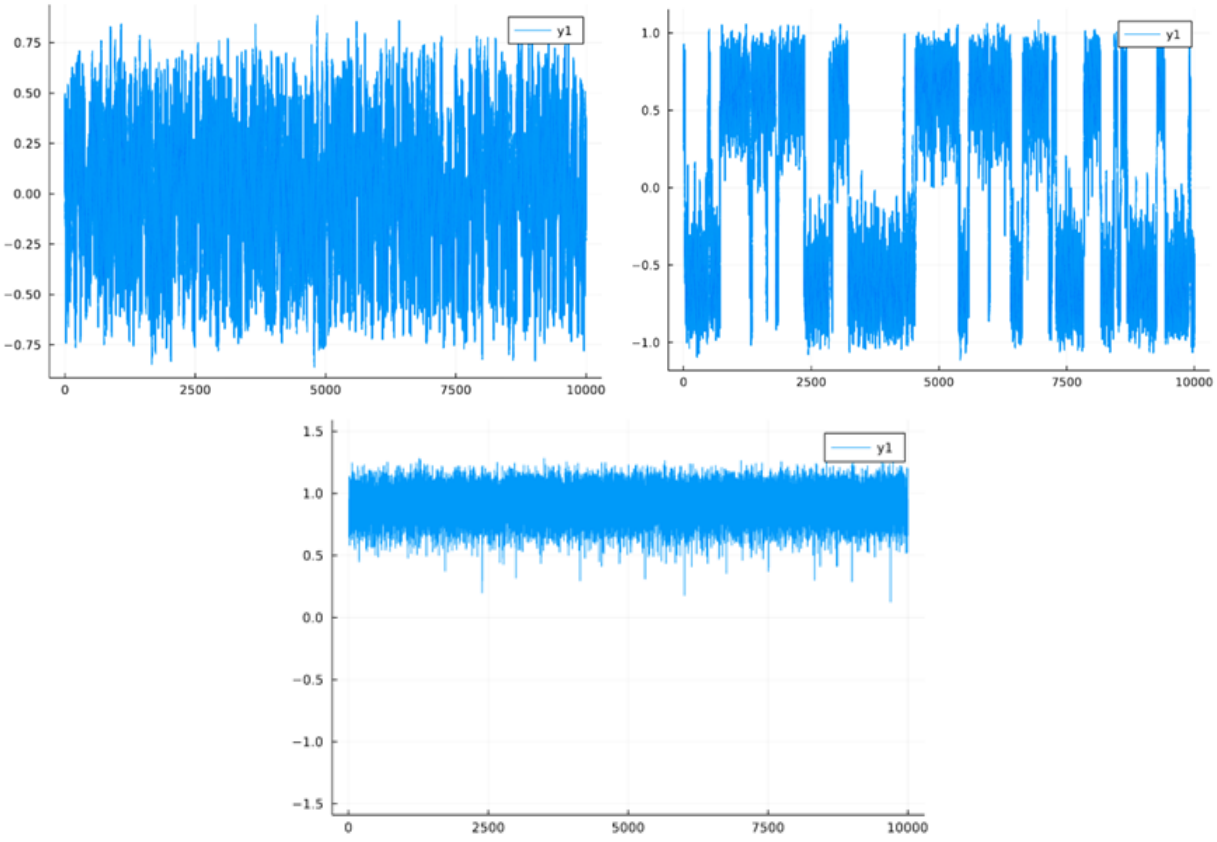


Figure 3: Output from generation code, with $a = 0, 0.45, 0.85$

Phase **I** clearly exhibits the stable noise behaviour expected. Phase **II** shows a sporadic alternation between the two points of stability on either side of the potential barrier. Lastly, in this specific example of a simulation at $a = 0.85$, phase **III** is 'stuck' or unable to leave the attraction of the right stable well, as the potential barrier is too large. The noise itself inside this well is also damped as a result of the strong recovery rates in

the potential.

When we examine the results over control parameter ranges, we can observe how the phase transitions appear on a large scale. This can be seen in Figure 4. The variance over the transition does not diverge as expected near 0, where phase **I** begins, instead we observe no discernible jump or divergence to infinity at all. Similarly for AC1, there are no indicators for the transition between phases **I** and **II**. This transition in the direction of decreasing control parameter can be considered a supercritical pitchfork bifurcation (Strogatz (2018)), as the two metastable wells and the unstable flat at the top of the potential barrier collapse to a single stable point at the origin.

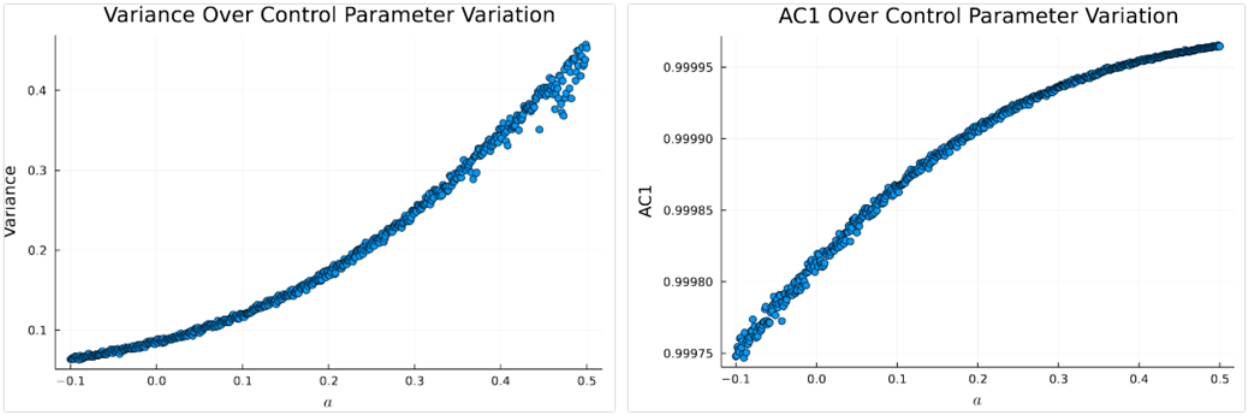


Figure 4: The transition between Phase **I** and **II**

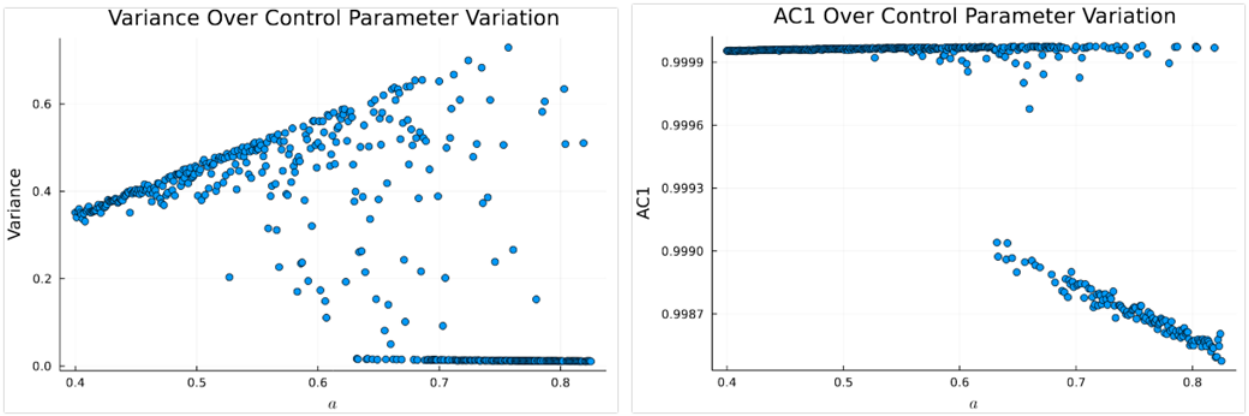


Figure 5: The transition between Phase **I** and **II**

Phase **II** to **III** is much clearer and can be seen in Figure 5. The variance of the simulation continues to increase as the control parameter increases- representing the dispersal of the wells horizontally, as well as the sharpening nature of the jumps. This linear increase disperses into a cloud of points, which meet near 0 for runs that initialise inside a well and remain there in the early stages of the system. The AC1 experiences a sharp drop, between runs that have their values heavily correlated and the phase **III** regime which sees the noise inside the well dominate over the forcing of the field on the potential. As expected, due to a lack of

averaging for runs, depending on initial condition and irrespective of the removal of transient behaviour, the transition can sometimes exhibit a jump even at high potential barrier levels if initialised close enough to the barrier. These outliers make it necessary to begin simulating multiple runs for a particular control parameter value selection and averaging the results. These results can be seen in Figure 6.

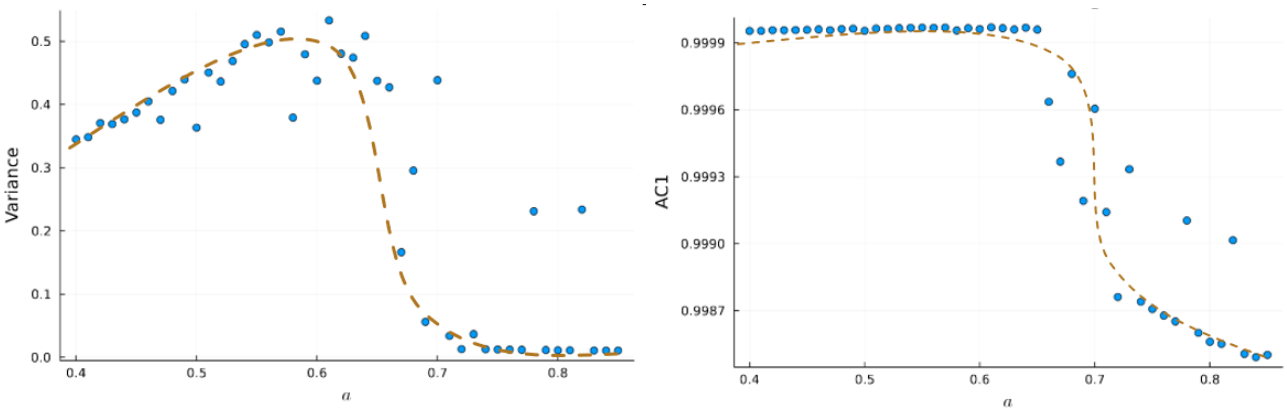


Figure 6: The averaged transition (over 3 runs) between Phase **I** and **II**, with an added tentative curve for clarity.

These averaged plots allow us to better quantify where the transition is likely to occur, and their results come from the averaging of 3 separate simulations. For our set parameter values, $a \approx 0.65$ is a good estimate for the second phase transition. We see a sporadic behaviour of variance on either side of the transition, but not a divergence to infinity. As we approach in the direction **II** $\longrightarrow$ **III** we can see that the variance diverges from its linear trend before collapsing rapidly to a lower regime where a linear trend continues. On this side, the variance, still has some outliers. It is expected these would flatten out were we to average more than 3 runs. The AC1 has the most robust behaviour under averaging, as it experienced less outliers (no cloud behaviour) like the variance. The shift between the two regimes is rigid, and we only see a few outliers in the direction **III** $\longrightarrow$ **II**, where prior to the transition, some values are removed from the generally increasing linear trend.

5.4 Hysterisis

In the second phase transition, **II** $\longrightarrow$ **III**, the cloud identified in the observed variance is a behaviour similar to that described in papers by authors such as Scheffer et al. (2012) in Figure 8. The observed delay between the cause (the increasing control parameter) and the effect (the phase transition) is what permits the system to be in both stable states simultaneously. What results is a curve that folds back on itself, where the system has a memory, or momentum. This curve is seen in Figure 7.

This behaviour was unexpected. Proper modelling and analysis is required, and a potential direction for future research.

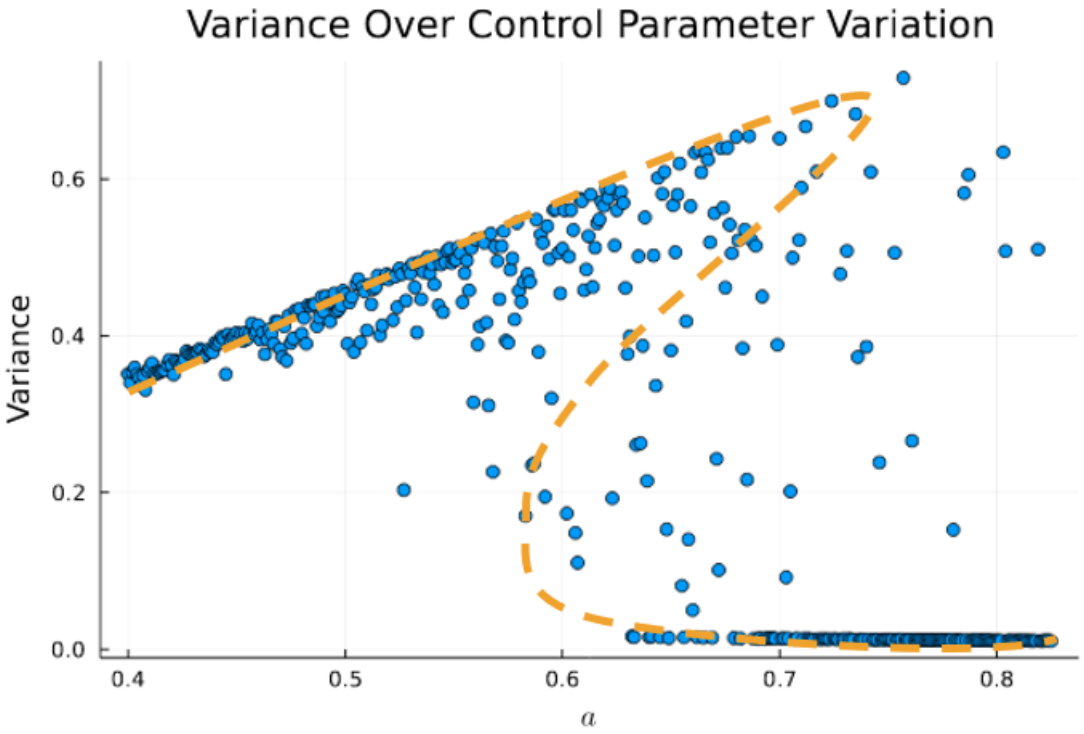


Figure 7: Observed hysterisis in the second phase transition, with hysteretic curve

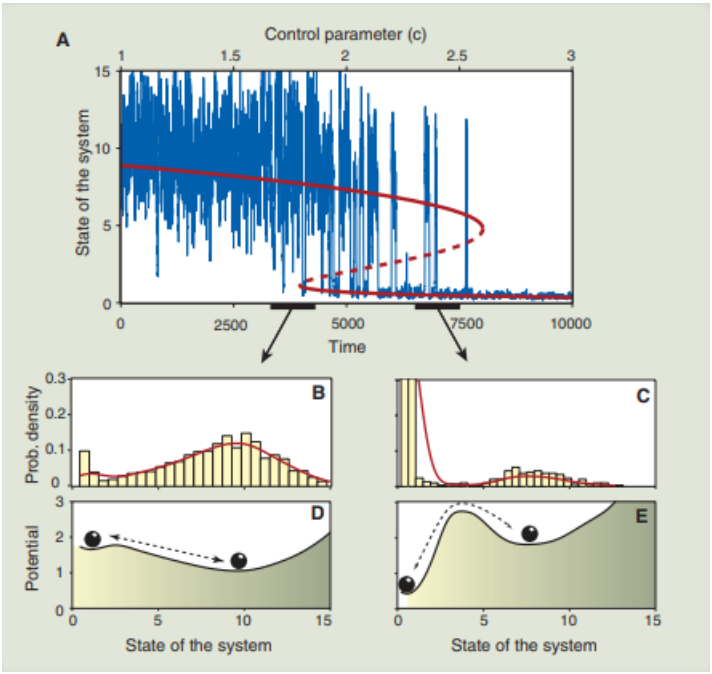


Figure 8: Taken from Scheffer et al. (2012), this figure shows an example of hysterisis in an SDE. Of note is the flickering to an alternate state which is similar to what we have seen in our results.

6 Further Applications

6.1 Lowered Resolution

A prominent application of time series/phase transition analysis is long form physical data.[16] The largest difficulty of which is that most methods of acquiring data result in a low resolution. Real-world systems evolve

continuously, our measurements are finitely precise, and can only happen at a discrete time interval. For that reason, it was considered constructive to consider how the phase transitions observed in our system changed under a decreased resolution. Instead of taking every data point available which is as precise as our time increment in the integration is, we integrate as normal, then artificially reduce the data set by observing at a set time interval. The two measures were retested for resolutions of 1 point in every 10,000, and 1 point in every 20,000. The results of the code in *Appendix 5* can be seen in Figures 9 and 10.

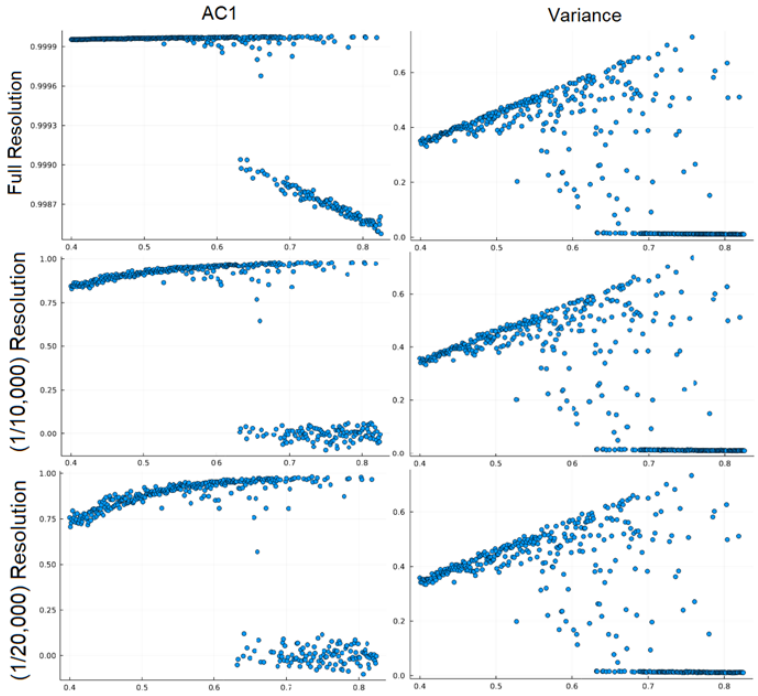


Figure 9: Results of a lowered resolution across the **II** $\longrightarrow$ **III** transition.

In Figure 9, we can see along the axes the immediate loss in information upon reducing the resolution. In auto correlation, the clear line in phase **II** very quickly becomes a curve with points more clouded out. The linear, decreasing regime in phase **III** upon a small information loss is immediately reduced to a random cloud around 0. This is the same behaviour as we would expect as the points are distributed nearly randomly, and the steep forcing of the field on the potential is no longer identifiable. Variance is far more resistant to the impact of a loss of resolution, and the largest impact visible is the linear trend in phase **II** prior to hysteretic behaviour having a larger spread of values compared to the full resolution data.

Figure 10 contains the same results but averages the values over 3 runs. For auto correlation, we can observe the decreased axis scale with any loss of resolution. As phase **II** approaches **I**, the auto correlation appears to be rapidly degrading. The linear decrease in phase **III** is also gone and replaced a simple 3-run analogue of the behaviour observed in 9. The variance is far more impacted in the 3-run average by resolution loss, as **III** remains quite robust but the transition **II** $\longrightarrow$ **III** becomes a cloud similar to the initial results in Figure 5. Such a selection of resolution appears to be susceptible to outlier behaviour- if measurements are taken at

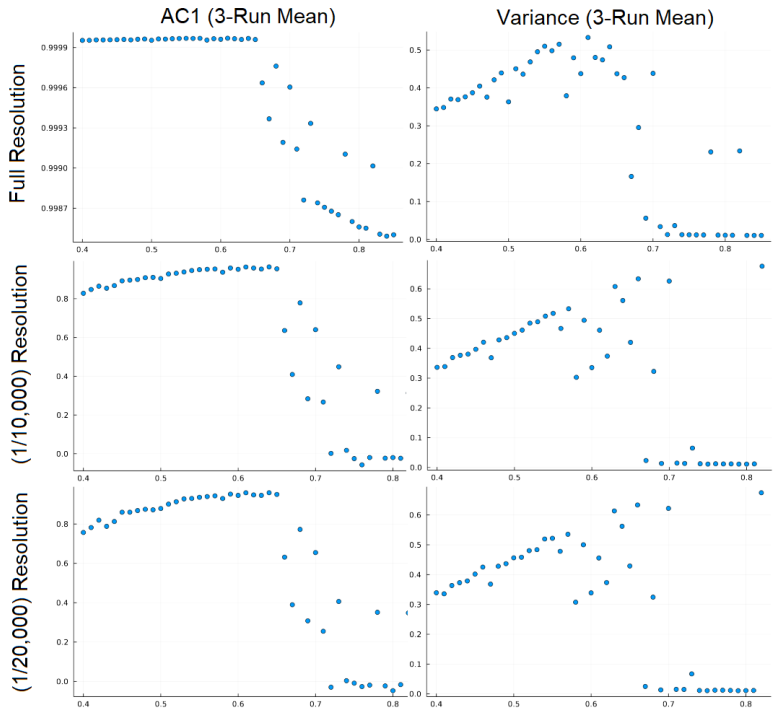


Figure 10: Results of a lowered resolution across the **II** $\longrightarrow$ **III** transition when runs are averaged (3 run average).

a particular time where noise/initialisation has created unique behaviour, our perceptions of upcoming phase transitions can be incorrect.

6.2 Ice Core Data

This analysis was applied to data from White et al. (2016) who had their data analysed by Garland et al. (2019) using an information theoretic approach. Our goal was to apply this elementary autocorrelative and variance approach and see if any climate transitions were identifiable. To clarify, this data set comprises of bored ice sheets which have had their levels of O^{18} and O^{16} measured, and taken as a ratio. This ratio is used as a temperature proxy for the surface at the time of measurement and is commonly referred to as $\delta^{18}O$. In our preliminary analysis, the focus was to identify the Younger and Older Dryas (YD and OD respectively) climate shifts, characterised by a significant return in the Northern Hemisphere to glacial conditions amongst a prevailing warming period. The time series of collected results generated from the code in *Appendix 6* is given in Figure 11.

There is a slight indication visible in Figure 12 of these climate events in autocorrelation, where both transitions are marked by a decrease prior, following by an increase. The variance does not provide any meaningful indication. To claim significant results when dealing with real world data, significance and associated testing is required, and thus further testing in this direction is an avenue for future research.

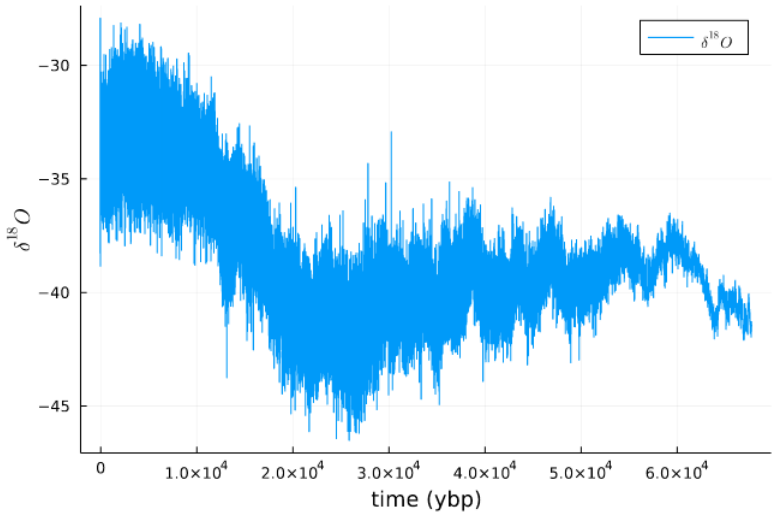


Figure 11: Ice Core data from White et al. (2016) plotted as a time series.

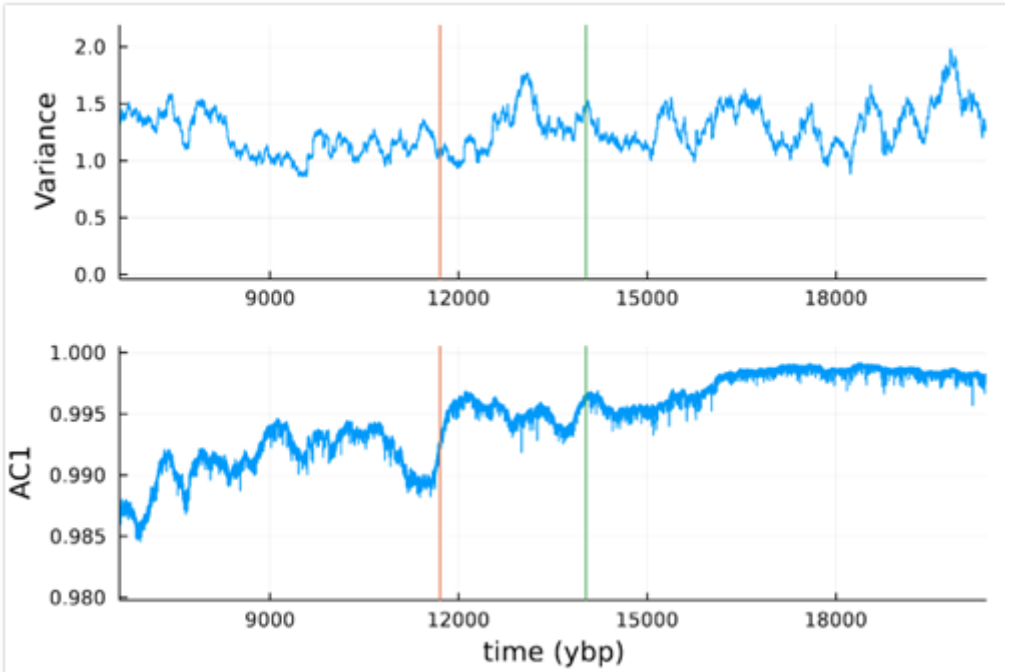


Figure 12: The result of a sliding function window along the ice-core time series. Window length = 5000. Younger Dryas indicated in orange, and Older Dryas indicated in green.

7 Conclusions

There are several effective methods established in the literature for analysing time series data. In our analysis, we aimed to characterise and identify phase transitions in a dynamical system from generated data. We were able to successfully identify from our reported measures significant and measurable behaviour changes in both variance and autocorrelation close to, and surrounding the second phase transition. These measures continued

to remain somewhat robust in the environment of a lower resolution time series. The analysis also proved capable of showing results in a real world data set. Several results remain open to further research. These include the appearance of hysteresis in the second phase transition, utilising measures such as rate of recovery and basin of attraction, and exploring the non-symmetric double well potential system.

References

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V. & Held, H. (2008), 'Slowing down as an early warning signal for abrupt climate change', *Proceedings of the National Academy of Sciences* **105**(38), 14308–14312.

Garland, J., Jones, T. R., Neuder, M., White, J. W. & Bradley, E. (2019), 'An information-theoretic approach to extracting climate signals from deep polar ice cores', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**(10), 101105.

Herzel, H. (1991), 'Risken, h., the fokker-planck-equation. methods of solution and applications. berlin etc., springer-verlag 1989. xiv, 472 pp., 95 figs., dm 98,—. isbn3-540-50498-2 (springer series in synergetics 18)'.

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., Pennell, P., Epstein, C., Bakay, R., Dichter, M. et al. (2001), 'Epileptic seizures may begin hours in advance of clinical onset: a report of five patients', *Neuron* **30**(1), 51–64.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M. & Sugihara, G. (2009), 'Early-warning signals for critical transitions', *Nature* **461**(7260), 53–59.

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van de Koppel, J., Van de Leemput, I. A., Levin, S. A., Van Nes, E. H. et al. (2012), 'Anticipating critical transitions', *science* **338**(6105), 344–348.

Strogatz, S. H. (2018), *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*, CRC press.

White, J., Alley, R., Archer, D., Barnosky, A., Dunlea, E., Foley, J., Fu, R., Holland, M., Lozier, M., Schmitt, J. et al. (2016), 'Stable isotopes of ice in the transition and glacial sections of the wais divide deep ice core', *US Antarctic Program (USAP) Data Center* .

8 Appendices

8.1 Appendix 1:

```
using Random, Distributions

#Wiener Increment, used to numerically integrate our noise forcing
function dW(delta_t,var)
    return sqrt(var*delta_t*2)*(rand(Normal(0,var))-rand(Normal(0,var)))
end
```

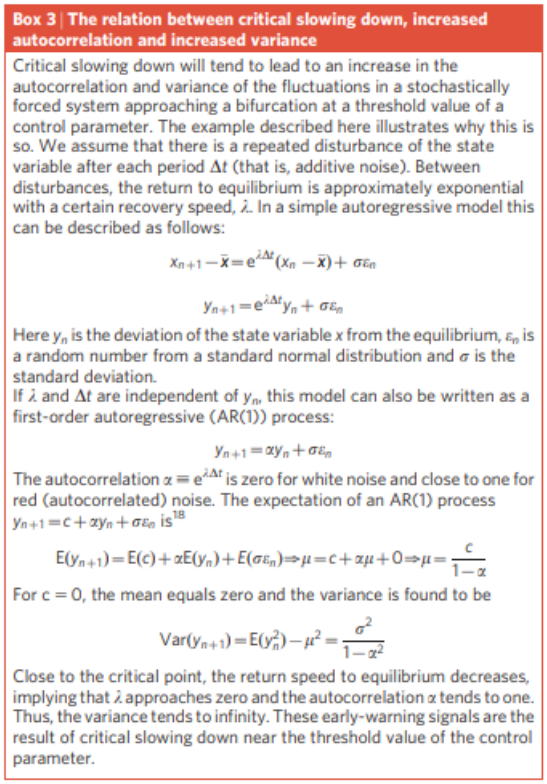
```
function simstart(t_end,dt,a,x_init,noise)
    t_init = 0
    #option for initialising the x-component
    x_init = rand(Uniform(-1,1))
    N = convert(Int64,((t_end - t_init) / dt))
    ts = collect(t_init:dt:(t_end))
    xs = zeros(N+2)
    xs[1] = x_init
    #the euler step, for integrating the system
    for i in 1:(size(ts,1))
        t = t_init + (i-1)*dt
        x = xs[i]
        xs[i+1] = x+dW(dt,noise)+dt*(a*x-x^3)
    end
    #we cut out transient behaviour of the function below, to ensure we see established systems
    xs = xs[10000:(N+1)]
    ts = ts[10000:(size(ts,1))]
    return([ts,xs])
end
```

8.2 Appendix 2:

```
using CSV, DataFrames

#The simulate function requires the simstart function

function Simulate()
    #the range and increment below dictates the resolution and control parameter range for the simulati
    for a in (A):0.01:(B)
        towrite = []
        points = simstart(10000,0.001,a,0,0.2)
        push!(towrite,Array(Float32.(points[1])))
        push!(towrite,Array(Float32.(points[2])))
        #results are written to the folder the code file is located in
        CSV.write("results$a.csv", DataFrame(towrite, :auto),
                    header = false)
```

```
    end
end


Simulate()
```

8.3 Appendix 3:



Box 3 | The relation between critical slowing down, increased autocorrelation and increased variance

Critical slowing down will tend to lead to an increase in the autocorrelation and variance of the fluctuations in a stochastically forced system approaching a bifurcation at a threshold value of a control parameter. The example described here illustrates why this is so. We assume that there is a repeated disturbance of the state variable after each period Δt (that is, additive noise). Between disturbances, the return to equilibrium is approximately exponential with a certain recovery speed, λ. In a simple autoregressive model this can be described as follows:

$$x_{n+1} - \bar{x} = e^{\lambda \Delta t}(x_n - \bar{x}) + \sigma \varepsilon_n$$

$$y_{n+1} = e^{\lambda \Delta t} y_n + \sigma \varepsilon_n$$

Here y_n is the deviation of the state variable x from the equilibrium, ε_n is a random number from a standard normal distribution and σ is the standard deviation.
If λ and Δt are independent of y_n, this model can also be written as a first-order autoregressive (AR(1)) process:

$$y_{n+1} = \alpha y_n + \sigma \varepsilon_n$$

The autocorrelation $\alpha = e^{\lambda \Delta t}$ is zero for white noise and close to one for red (autocorrelated) noise. The expectation of an AR(1) process $y_{n+1} = c + \alpha y_n + \sigma \varepsilon_n$ is[18]

$$E(y_{n+1}) = E(c) + \alpha E(y_n) + E(\sigma \varepsilon_n) \Rightarrow \mu = c + \alpha \mu + 0 \Rightarrow \mu = \frac{c}{1-\alpha}$$

For $c = 0$, the mean equals zero and the variance is found to be

$$\mathrm{Var}(y_{n+1}) = E(y_n^2) - \mu^2 = \frac{\sigma^2}{1-\alpha^2}$$

Close to the critical point, the return speed to equilibrium decreases, implying that λ approaches zero and the autocorrelation α tends to one. Thus, the variance tends to infinity. These early-warning signals are the result of critical slowing down near the threshold value of the control parameter.

Taken from box 3 in Scheffer et al. (2009).

8.4 Appendix 4:

```
using StatsBase, CSV, DataFrames, LaTeXStrings, Statistics, Plots
autol1 = []
#the two append functions can be commented out to switch between autocorrelation and variance calculati


for a in (A):0.001:(B)
    file = CSV.read("(INSERT RESULT PATH HERE)\\results$a.csv",DataFrame, header=0)
    #append!(autol1,autocor(file[:,:2], [1]))
    #append!(autol1,var(file[:,:2]))
end
scatter(Array((A):0.001:(B)),autol1, legend=false)
```

8.5 Appendix 5:

```
# A and B here are control parameter ranges, and C in the append functions represents the frequency wit
autol = []
for a in (A):0.001:(B)
    file = CSV.read("(INSERT RESULT PATH HERE)\\results$a.csv",DataFrame, header=0)
    append!(autol,autocor(file[:,:2][begin:C:end], [1]))
    #append!(autol,var(file[:,:2][begin:C:end]))
end
```

8.6 Appendix 6:

```
using Random, Distributions, Plots, RollingFunctions, StatsBase, CSV, DataFrames, LaTeXStrings, Statist

file = CSV.read("(ICE CORE DATA PATH)",DataFrame, header=0, delim=' ')
file = file[:,[1,3]]
points = file[:,:]

#the below function shifts a time series by one value and returns both the shifted and original list
function shift1(array)
    temp1 = copy(array)
    temp2 = copy(array)
    popfirst!(temp1)
    pop!(temp2)
    return([temp1,temp2])
end

points1 = Array(points)[:,1]
points2 = Array(points)[:,2]

# the below two functions make sure to demean windows as we roll over the data
function demeanvar(array)
    m = mean(array)
    f(x) = x-m
    return(var(f.(array)))
end

function ac1(array)
```

```
    return(autocor(array,[1],demean=true)[1])
end

#W in the below functions represents the sliding window length. We used W=5000

time = rollmean(points1,W)
variances = rolling(demeanvar, points2, W)
ac1s = rolling(ac1, points2, W)
```