nature portfolio

Peer Review File

Network representations of attractors for change point detection

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this study, the authors propose an ingenious approach for online and supervised change point detection. The method, termed "attractor network," synergizes concepts from dynamical systems (particularly phase space reconstruction) and complex networks and can be regarded as a three-step process:

- First, the authors employ time delay embedding on the training data to estimate the system's attractor.

- Second, they construct a sparse point cloud representation (spatial network) of the attractor, which they subsequently discretize to form a dynamics network based on transition frequencies among points in the discretized phase space.

- Lastly, the authors utilize the testing data to identify change points. Their method hinges on the notion that observing improbable transitions or transitions absent in the training data signifies shifts in the system's operational regime. This task involves computing a properly normalized entropy linked to transition probabilities between each pair of nodes in the dynamics network.

The primary concepts of this method are presented in the main text, with exhaustive details relegated to the Methods section. Notably, the authors furnish two main applications of their "attractor network": firstly, detecting the onset of ventricular fibrillation in ECG signals, and secondly, addressing in silico time series related to the Chua chaotic oscillator and its surrogates. The results not only showcase the effectiveness of the method in pinpointing change points within time series but also underscore its superiority over techniques grounded in simple sliding window statistics and permutation entropy.

This manuscript is dense and very well written. The proposed approach holds significant promise and is poised to contribute substantially to a broad spectrum of change point detection applications. Nonetheless, there are some points that I wish the authors to consider before my final recommendation.

1) In the introduction, when discussing the categorization of "change point detection methods," I encourage the authors to provide further contextualization and references related to the various types, particularly emphasizing unsupervised methods, which are only briefly mentioned.

2) The foundational concepts underpinning the proposed approach closely resemble techniques within topological data analysis (TDA). Considering the existence of multiple TDA methods dedicated to anomaly detection, I recommend establishing a connection or possibly drawing comparisons between these methods to situate the current work more effectively within the existing literature.

3) While the ingenuity of the proposed approach is undeniable, it could potentially appear daunting to researchers from diverse fields, such as Biology and Medicine. Although the authors have made their code available for result replication, its accessibility could be enhanced. In this regard, I suggest the authors contemplate developing a well-documented software package for their method. I acknowledge that this endeavor might require time, and therefore, consider this only as a suggestion for the future. However, I am confident that such an initiative could significantly amplify the impact of the "attractor network."

4) The authors touch upon the challenge of defining "false positives." Nevertheless, upon examining Figure 5, one might argue that all detections preceding the ventricular fibrillation flag could be construed as false positives. The authors assert that this assumption cannot be made due to the

"nature of the data." While this standpoint is debatable, I wonder if a similar pattern would emerge in the analysis of artificial time series.

Misprints:
- "plotting the *the* trajectory"
- "chaotic *oscilllator*"
- "ECG recordings *are* contain complex"
- "pseudoperiodicity can be *attribute* to minor"
- "the surprise for transitions with missing values *were* set
to 0"
- "are used *to* control the density"
- "as *an* detected change"
- "One of the most commonly employed methods of reconstructing the phase space of a system from observed scalar time series *via* the method of time"
- "requires *the* solving the non-trivial"
- "in cycles until the *the*"
- "required in *the* the grid-based"
- "This *is* corresponds to"
- "may otherwise be *difficulty* to detect"
- All references to figures in the appendix are pointing to figures in the main text.

Reviewer #3 (Remarks to the Author):

The work titled "Network representations of attractors for change point detection" study the important and relevant question about change point detection, or in other words, the moment where can be observed structural changes that are influential for dynamical events. They use analysis of surprise from phase space reconstruction from data, which can be challenging depending on several parameters and methods, and apply into ECG data and other chaotic systems. They also compare with some other literature methods with good results. The work is quite clear, well written and present a deep bibliography analysis. I do think it is of the interest of the reader of this journal.

Reviewer #4 (Remarks to the Author):

The authors are proposing a statistic based on a kind of transition matrix for detecting change points from a time series. As far as I have read their manuscript, the proposed statistics called "surprise", which is related to some entropy, seems novel and shows its good performance in Fig. 7. Therefore, the proposed method should be published in an appropriate form. However, I cannot recommend the current manuscript as it is because I have several concerns on the manuscript.

a) First of all, I could not understand how a spatial network is constructed. I will say that the proposed method seems incomplete. Namely, I could not interpret the last line of Step 5 to renew the set A. What are the authors doing in this line? To clarify this point, please define the backslash as well as the " ' " explicitly so that every reader can interpret this line of the mathematical expression in the same way. In addition, is it always possible to have k_{max} less than 3? If so, why is it possible? Moreover, readers would find much easier if the authors explain consicely what each of their steps does.
b) I could not understand the intuitive interpretation of the statistic "surprise". This statistic looks like the conditional entropy, but these absolute values look different. Why do the authors want to define this statistic in this way?
c) Is the statistic "surprise" an invariant measure?
d) As far as I have searched in the existing literature, there are some other methods of change point

detections in the context of nonlinear time series analysis. Thus, please provide the comparisons of the proposed method with such methods in addition to windowed permutation entropy and the methods assuming the linearity for the underlying dynamics. (Maybe, the authors should have to review such methods.)

e) Is the time interval appropriate for obtaining the moving permutation entropy in Fig. 7? (If we do not have to choose the delay in the proposed method, then it would be their merit. But, because the authors seem to use the discretized phase space, they are possibly using delay coordinates, and thus the authors might have to choose some delays, I guess.)

f) Why do the authors want to use AAFT surrogates in this context? If I were among the authors, I would be interested in differentiating two different regimes in nonlinear dynamics.

g) Before applying the proposed method to a real dataset, please examine the performance of the proposed method with some toy models with more depths. The results shown in Fig. B6 do not appeal to me much because one may use MSTD or MPE if their performances there are correct. Furthermore, the results shown in Fig. C7 do not make sense to me: if the proposed method quantifies the complexity for the underlying dynamics, please name the statistic more appropriately. If the authors want to call the statistic "surprise", what surprise did the authors find from the panel a) of Fig. C7?

These are my major concerns.

I also have some minor concerns.

1. Line 444, SToPS needs more explanations since it has not become a common practice and the average readers have to read Ref. [52] to understand the meanings of the current manuscript perfectly.
2. Line 738, "algorithm" seems "the algorithms".
3. Line 763, "network-base" seems "network-based".
4. Line 806, "point" seems "points".
5. As for the right hand sides of Eqs. (7) & (8), probably the authors have forgotten to make them some norms.
6. Line 875, "maybe" seems "may be".
7. Line 933, the right hand side needs to depend on t, which is missing.
8. Line 1402 or 1403, ".." seems "."
9. Line 1413 "is also tracked" seems "are also tracked".

Review Response Letter

Dear Communication Physics Editorial Board and Reviewers,

We would like to thank the editor and reviewers for taking the time to consider our submission and providing valuable comments. For our resubmission, we have provided in this document a point-by-point response to the concerns raised by the reviewers. Reviewer comments are restated in blue and followed by a response in red. Where possible, we have included in-text citations to changes made in the manuscript in italics, and have also highlighted them in red in the main manuscript.

Reviewer 1

In this study, the authors propose an ingenious approach for online and supervised change point detection. The method, termed "attractor network," synergizes concepts from dynamical systems (particularly phase space reconstruction) and complex networks and can be regarded as a three-step process:

- First, the authors employ time delay embedding on the training data to estimate the system's attractor.
- Second, they construct a sparse point cloud representation (spatial network) of the attractor, which they subsequently discretize to form a dynamics network based on transition frequencies among points in the discretized phase space.
- Lastly, the authors utilize the testing data to identify change points. Their method hinges on the notion that observing improbable transitions or transitions absent in the training data signifies shifts in the system's operational regime. This task involves computing a properly normalized entropy linked to transition probabilities between each pair of nodes in the dynamics network.

The primary concepts of this method are presented in the main text, with exhaustive details relegated to the Methods section. Notably, the authors furnish two main applications of their "attractor network": firstly, detecting the onset of ventricular fibrillation in ECG signals, and secondly, addressing in silico time series related to the Chua chaotic oscillator and its surrogates. The results not only showcase the effectiveness of the method in pinpointing change points within time series but also underscore its superiority over techniques grounded in simple sliding window statistics and permutation entropy.

This manuscript is dense and very well written. The proposed approach holds significant promise and is poised to contribute substantially to a broad spectrum of change point detection applications. Nonetheless, there are some points that I wish the authors to consider before my final recommendation.

Thank you very much for the kind comments and positive response to our submitted work. We are happy and encouraged to hear that our work holds potential in providing a significant contribution to the field.

Comments:

In the introduction, when discussing the categorization of "change point detection methods," I encourage the authors to provide further contextualization and references related to the various types, particularly emphasizing unsupervised methods, which are only briefly mentioned.

Thank you for highlighting this deficit. We agree that direct discussion regarding the unsupervised methods is not clear. In fact, the discussion of decision trees, sliding windows and Bayesian approach (references provided) in the original submission were actually examples of techniques used to achieve unsupervised methods. We have edited this section in the introduction and have restructured the text to highlight this point and provide a better context for this method.

The second classification on supervision describes whether ground truth labels are known a priori. Supervised methods construct a reference model based on a pre-defined ground truth (e.g user provided labels of normal vs. abnormal). Here, we use the term 'model' loosely to describe any constructed system, set of statistics or parameters that characterises the ground truth state. Deviations from the model are then used to infer system changes.

In contrast, unsupervised methods attempt to circumvent the requirement of a pre-defined ground truth by comparing new incoming observations against recently observed data. This is usually achieved by comparing statistics (mean, *probability density, permutation entropy) of data between two moving windows (subsequences) in the time series separated by some lag. A change point is flagged if the statistics of the most recent observed window greatly differ from prior observations. Some methods used for change point detection include decision trees, support vector machines, and statistical approaches such as Gaussian mixed models, Gaussian processes and Bayesian methods. However, the usage of statistical measures typically relies on the data adhering to some stationary distribution. This does not account for temporal dependencies often present in dynamical systems, which may be useful in uncovering and characterising changes in the underlying data generating process. For the scope of this paper, we will focus specifically on an online supervised change point detection method.*

The foundational concepts underpinning the proposed approach closely resemble techniques within topological data analysis (TDA). Considering the existence of multiple TDA methods dedicated to anomaly detection, I recommend establishing a connection or possibly drawing comparisons between these methods to situate the current work more effectively within the existing literature.

We thank the reviewer for this comment and suggestion and have included a brief discussion on some recent development of TDA based methods for the purpose of change point detection in the introduction. We hope this provides a better context for the position of our work in the wider literature of change point detection.

The use of a geometrical approach has also inspired the application of various topological data analysis (TDA) methods for system characterisation and change point detection. Persistent homology has previously been found to be useful in characterising the state of dynamical systems, and also experimentally for classifying breathing signals and chatter detection. More recently, persistent homology methods have also been adapted for the purpose of change point detection from temporal data. For a sequence of point clouds, mappings of Betti sequences and Wasserstein distances between consecutive persistence diagrams have been used as measures for identifying change points. Building on this, persistence diagram-based change-point detection (PERCEPT) utilising ℓ_2 divergences between persistence histograms was proposed as an improvement on previous methods and was used to analysing solar flare images. However, whilst sequences of points clouds can be constructed from scalar time series data (e.g. via sliding windows), systems whose attractors have highly complex geometry may require large windows in order for the underlying geometry to be well captured by persistent homology. This consideration in conjunction with the the poor computational scaling of persistent homology algorithms, and the quick detection response time typically desired in online change point detection tasks can make direct application of these methods challenging.

While the ingenuity of the proposed approach is undeniable, it could potentially appear daunting to researchers from diverse fields, such as Biology and Medicine. Although the authors have made their code available for result replication, its accessibility could be enhanced. In this regard, I suggest the authors contemplate developing a well-documented software package for their method. I acknowledge that this endeavor might require time, and therefore, consider this only as a suggestion for the future. However, I am confident that such an initiative could significantly amplify the impact of the "attractor network."

Thank you for this suggestion. The original code has been prepared in a notebook style format such that interested users should be able to edit the inputs such that it can be applied to another data set. However, we understand that this may be daunting for researchers in other fields that may not be used to manipulating data in this way. We fully agree with the reviewer and have plans of coding up this method as a usable package for the public in the very near future as most of the algorithm groundwork is already present.

The authors touch upon the challenge of defining "false positives." Nevertheless, upon examining Figure 5, one might argue that all detections preceding the ventricular fibrillation flag could be construed as false positives. The authors assert that this assumption cannot be made due to the "nature of the data." While this standpoint is debatable, I wonder if a similar pattern would emerge in the analysis of artificial time series.

In our discussion, we can broadly identify three main sources of "false positives". Hopefully the following will provide some clarifications.

(1) False positives due to missing values in the data

As our method tries identify changes in the dynamics/vector field in phase space, missing data (that has been set to 0) would easily be considered as anomalies as the trajectory will be seen as jumping to the origin, which would be construed as odd behaviour. In application, if the data dropouts are intermittent for short periods, one may consider several interpolation methods (linear, nonlinear or predictive) to try and "fill in" the missing data. However, for simplicity, we have replaced these missing values with 0s.

(2) False positives for the ECG after the first VF episode

The argument we present in the text regarding the difficulty of defining false positives would primarily apply to this case i.e. after VF, there may transient stress to the heart which can fundamentally change the "healthy/normal" dynamics of the heart. Therefore, the previous learned attractor network may not longer be an accurate baseline from which to detect anomalies. However, we agree with the reviewer that this likely does not apply to cases where detections are flagged prior to the first VF, and have clarified this in text accordingly in Section 2.1.

(3) False positives prior to VF

This is likely the case that the reviewer has pointed out that is of significant interest. The first type of this case to consider is illustrated in Section 4.5.2, patient cu04. In this situation, a detection is flagged briefly for a time period prior to the onset of VF, but does not actually trigger at the exact annotated time of VF in dataset. In this case, the earlier flag, which may be seen as a "false positive", is actually indicative of a pre-emptive detection.

However, in the other cases for false positives where there are short term occasional flags by the algorithm, this largely boils down to threshold selection for flagging positives. A majority of the observed false positives are typically short-lived, and from the ECG case, appears to be the characteristic differentiator between a true and false flag. Real VF tends to produce sustained positives. For clarification in Section 2.1, we have briefly expanded on false positives prior to the first VF detection, and their relevance to persistent positive flags.

We note, that false positives can and do occur for times exceeding more than 5 seconds prior to the annotated onset of VF. However, from Figure 2.1 these flags are relatively brief and are characteristically different to the persistent flags corresponding to a true VF episode.

When applying a sliding window approach, the type of measure used to identify anomalies (moving statistics, permutation entropy, surprise) is the main determinant of how well these false positive are avoided. Generally speaking, good measures should have statistics or magnitudes that clearly separate normal and abnormal, which would make determining the threshold cutoff clearer and consequently lend themselves more effective when applied with sliding window approach.

Practically, apart from the choice of measure, there is also a compromise of identifying false positives and negatives, which is usually influenced by hyperparameters such as the window length, gap between windows, etc. One way to tackle these false positives would be to adjust window length, threshold cutoffs and hyperparameters. Another alternative approach would be to apply an additional moving average filter on top of the existing binary outputs to ensure that only persistent flags (i.e. real VF onsets) get flagged. However, both these approaches may come at a cost for the time taken for a real change to be detected, potentially sacrificing the benefits of pre-emptive detection.

The novelty of our work lies in the use of a "surprise" entropy measure that better incorporates dynamical information into the resulting moving statistic which may perform better alongside the moving window approach, compared to other statistical measures.

As noted by the reviewer, most of the above issues are not encountered to the same degree (if at all) when analysing artificial time series. As demonstrated Figure 7 (Amplitude Adjusted Fourier Surrogate for Chua) and Figure B6 (Phase coherent vs Non phase coherent Rossler), the occurrence false positives are much lower. In these cases, the underlying vector field for the normal regime is static, and any deviations are quickly picked up by the surprise. This contrasts with experimental data such as those of ECG, where the underlying dynamics in the healthy regime may not be perfectly stationary.

Reviewer 3

The work titled "Network representations of attractors for change point detection" study the important and relevant question about change point detection, or in other words, the moment where can be observed structural changes that are influential for dynamical events. They use analysis of surprise from phase space reconstruction from data, which can be challenging depending on several parameters and methods, and apply into ECG data and other chaotic systems. They also compare with some other literature methods with good results. The work is quite clear, well written and present a deep bibliography analysis. I do think it is of the interest of the reader of this journal.

Thank you for the positive comments and time taken by the reviewer to consider our work.

Reviewer 4

The authors are proposing a statistic based on a kind of transition matrix for detecting change points from a time series. As far as I have read their manuscript, the proposed statistics called "surprise", which is related to some entropy, seems novel and shows its good performance in Fig. 7. Therefore, the proposed method should be published in an appropriate form. However, I cannot recommend the current manuscript as it is because I have several concerns on the manuscript.

Comments:

First of all, I could not understand how a spatial network is constructed. I will say that the proposed method seems incomplete. Namely, I could not interpret the last line of Step 5 to renew the set A. What are the authors doing in this line? To clarify this point, please define the backslash as well as the " ' " explicitly so that every reader can interpret this line of the mathematical expression in the same way. In addition, is it always possible to have k_{max} less than 3? If so, why is it possible? Moreover, readers would find much easier if the authors explain concisely what each of their steps does.

Thank you for the feedback regarding the algorithms. The first two paragraphs of Section 4.2.1 elaborate on the general process of the algorithm. To improve readability, we have broken up the previous Step 5 into multiple stages and clarified the mentioned variables alongside some mistypes. Hopefully these improves the readability of the method.

To expand on the actual approach, the construction of the spatial network involves gradually including batches of candidate points (in embedded space) into the spatial network, followed by an algorithm to thin out the point cloud such that the density of points is bounded by a size scale ϵ. This is done to ensure that the algorithm is parsimonious and scales well with dimension. The thinning process is achieved by replacing node clusters (i.e. regions of high density) with an average centre of mass.

This approach of thinning starts from the highest possible degree and progressively replaces clusters with single points until no triangles (i.e. $k_{max} = 3$) remain. Because in each step, the number of points removed will always be greater or equal to the number of points added, this will result in a sparsification of the points in the spatial network, and thus will always be able to reach $k_{max} < 3$. We have added a brief statement at the end of Section 4.2.1 to clarify this.

I could not understand the intuitive interpretation of the statistic "surprise". This statistic looks like the conditional entropy, but these absolute values look different. Why do the authors want to define this statistic in this way?

One way to relate surprise to entropy would be that entropy is the expected level of surprise averaged across all possible outcomes. The reviewer is correct in noting that Equation 12 is very similar to surprise. If we define p_{ij} as the probability of a transition from node i to node j and $y_t \in \{1, 2, ..., N\}$ as the index of the current closest node at a given time, this can be written as a conditional probability,

$$p_{ij} = P(y_{t+1} = j \mid y_t = i).$$

We may define the conditional entropy H_{ij} as

$$H = -\sum_{i,j \in \{1,\dots,N\}} p_{ij} \log(p_{ij}).$$

However, the conditional entropy is a measure of the expected level of surprise across all possible pairs of transitions $i \to j$, whereas Equation 12,

$$S = -\eta_i \log(p_{ij}) = -\frac{H_{max}(i)}{H(i)} \log(p_{ij}),$$

is instead a measure of the instantaneous surprise of a given transition $i \to j$ weighted by a normalising constant η_i which accounts for how informative a given transition is based on the degree and transition probabilities of its origin node. We have amended the paragraph preceding Equation 12 to remove the term "entropy" to avoid this potential confusion.

For example, consider an origin node i_1 with outdegree 10, each with equal probabilities compared against another node i_2 with outdegree 3. The unnormalised surprise would provide yield the following values:

$$S_{i_1} = \log(\frac{1}{10}) = 2.3 \quad (1)$$

$$S_{i_2} = \log(\frac{1}{3}) = 1.09 \quad (2)$$

Since in both cases all possible outcomes across the set of available outward links are equally likely, the surprise of both scenarios should be equal. However, the higher outdegree of i_1 results in an inflated value of the surprise. Additionally, it can also be argued that observed transitions from nodes with outward transition probabilities that are not uniform should be seen as more informative/significant. Both of this factors are accounted for with normalising constant η_i.

Is the statistic "surprise" an invariant measure?

The surprise measure is the result of a comparison between and observed and an expected trajectory. Therefore, surprise values are dependent on input test observations and thus would not be an invariant measure. However while surprise is not an invariant measure, we note that true invariant measures tend not to always be useful indicators of change points as they often require large amounts of data to estimate well

As far as I have searched in the existing literature, there are some other methods of change point detections in the context of nonlinear time series analysis. Thus, please provide the comparisons of the proposed method with such methods in addition to windowed permutation entropy and the methods assuming the linearity for the underlying dynamics. (Maybe, the authors should have to review such methods.)

In response to the reviewer, we have included in the Appendix D additional comparisons between the attractor network approach and two recently proposed nonlinear change point detection methods, quadrant scan (QS) and modularity scan (MS), based on recurrence quantification analysis (RQA). We understand whilst there may also be other nonlinear methods, we have chosen these comparison methods for several reasons: (1) their conceptual simplicity and ease of implementation, and (2) the similarity they share with the attractor network approach in approximating changes in the vector field.

In our additional tests, we find that both RQA methods perform comparably against the attractor networks. This is unsurprising as the all methods rely on quantifying and detecting changes (before and after some reference point) in the underlying vector field of the system.

Is the time interval appropriate for obtaining the moving permutation entropy in Fig. 7? (If we do not have to choose the delay in the proposed method, then it would be their merit. But, because the authors seem to use the discretized phase space, they are possibly using delay coordinates, and thus the authors might have to choose some delays, I guess.)

For the analyses of the artificial systems in Figure 7, we calculate the permutation entropy based on an embedding dimension of 4 and the lag taken as the first selected lag from the SToPS algorithm. Whilst we recognise that the surprise method using only an embedding dimension of only 3, this dimension would only result in a total of 6 possible symbols, which would not be sufficient to adequately estimate the permutation entropy. Additionally, the permutation entropy approach uses uniform lags.

As for the selection of lags, the permutation entropy was calculated using uniform lag τ selected using SToPS. Because the attractor network uses a non-uniform embedding approach, we have conducted additional tests with attractor network to verify its performance even with uniform embedding.
From Figures 1 and 2, we find that the attractor network is still able to outperform the permutation entropy approach even when using uniform embedding. However, non-uniform embedding is used in our analyses as it allows for dealing with dynamics with multiple time scales and can provide a better reconstruction of the underlying attractor. Due to the length limits of the manuscript, we have not included Figures 1 and 2 within our manuscript. However, we hope

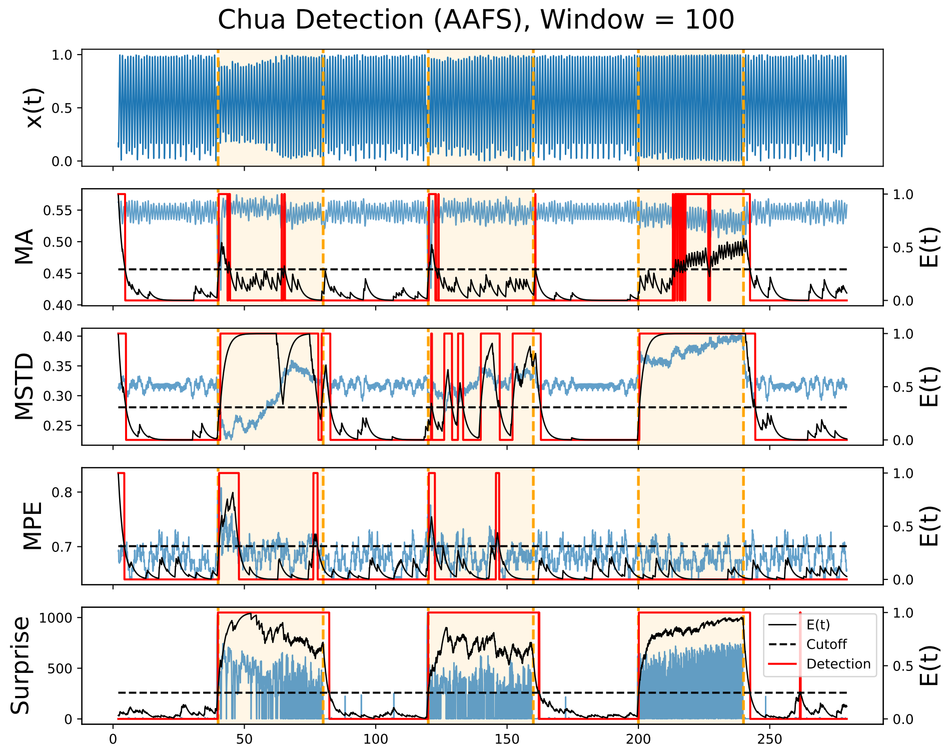

Figure 1: Chua AAFS test results with uniform embedding with first lag from SToPS

Figure 2: Rössler PC vs. NPC test results with uniform embedding with first lag from SToPS

the results help illustrate our points.

Why do the authors want to use AAFT surrogates in this context? If I were among the authors, I would be interested in differentiating two different regimes in nonlinear dynamics.

Iterated AAFT surrogates were chosen as the change point detection task due to its difficulty. Generally, iterated AAFT surrogates preserve the overall statistics (Mean and STD), as well as the Fourier power spectrum of a given time series. Applied to any set of time series, the result is an output time series that preserves the previously mentioned quantities, but no longer possess the same dynamics. Therefore, the AAFT surrogates test should in theory assess whether a given method is able to distinguish between changes in the underlying dynamics.

However, AAFT surrogates are not able to preserve the moving statistics of a time series. Additionally, iterated AAFT surrogates of most chaotic systems generally produce time series that are visually quite different from the original time series (e.g. AAFT of a Lorenz time series). However, we select the Chua oscillator operating in the single scroll regime as it exhibits relatively regular, albeit chaotic oscillatory dynamics with AAFT surrogates that also look visually quite similar. The result is a time series whose moving statistics (average and standard deviation) can appear quite similar to the original time series, but no longer possess the same dynamics.

We also would like to highlight that the analysis between two different regimes of nonlinear dynamics are also demonstrated in our paper in two different experiments in the Appendix: (1) Phase-coherent (PC) vs non-phase coherent (NPC) Rossler, and (2) Single-scroll vs. double scroll Chua.

In the first instance, the artificial time series consists of a system that undergoes bifurcations between the phase-coherent and non-phase coherent regime of chaotic dynamics. This example was taken to demonstrate that the attractor network approaches leverages on anomalies in the underlying attractor to identify change points. Based on Fig B5, the PC Rossler's attractor is spatially close to a subset of NPC attractor, and so it would be expected than an attractor trained on PC would be able to detect changes to NPC, but not the converse. This task is demonstrated in Figure B6 where the attractor network outperforms the other baseline methods. We also checked the converse (NPC attractor network to detect PC) and verified that it does not work in this case due to space of the NPC attractor effectively already containing the PC attractor.

The purpose of the second test, is to identify if the surprise metric is able to detect gradual bifurcation changes (as opposed to sudden changes) which suggest that it may be used as a rough measure of far a system has deviated from its "normal" regime. The results and discussion of this is provided in Appendix C.

Before applying the proposed method to a real dataset, please examine the performance of the proposed method with some toy models with more depths. The results shown in Fig. B6 do not appeal to me much because one may use MSTD or MPE if their performances there are correct. Furthermore, the results shown in Fig. C7 do not make sense to me: if the proposed method quantifies the complexity for the underlying dynamics, please name the statistic more appropriately. If the authors want to call the statistic "surprise", what surprise did the authors find from the panel a) of Fig. C7?

Each of the toy model tests presented were aimed at assessing a specific aspect of the attractor network algorithm. The primary of this is the AAFS surrogates, a situation where time series appears visually similar moving statistics (e.g. moving average) may not be informative. The additional cases of toy models in the Appendix sections B and C are aimed at verifying two other behaviours of the attractor networks algorithm.

For section B, the aim of detecting changes between PC and NPC is to verify that the attractor network method indeed works by detecting changes in the underlying attractor and vector field. This is demonstrated by its ability to detect NPC when trained on PC, but not the other way around. We have edited this section to clarify this.

From Figure B5 the PC attractor is close to a subset of the NPC regime. Hence, the attractor network approach that quantifies attractor changes would be able to detect changes from PC to NPC but not the reverse. This feature was used to test the attractor network approach in detecting changes in attractor structure in phase space

For section C, the aim of the test was to see if surprise metric had the potential to also quantitatively estimate the degree of change from its normal regime. In Figure C7, whilst it is plotted above a bifurcation diagram, the value

$\bar{S}(\alpha)$ is not presented as a measure of complexity, but rather as a measure of degree of change of the system from its 'normal' regime. We have highlighted in the text:

The relatively gradual increase in surprise with increasing perturbation of the bifurcation parameter suggests that the attractor network may be sensitive in measuring the magnitude of change in a given system.

I also have some minor concerns. Line 444, SToPS needs more explanations since it has not become a common practice and the average readers have to read Ref. [52] to understand the meanings of the current manuscript perfectly.

We have expanded on basic outline of SToPs in the paper to provide readers a bit more context for how the method works.

The SToPS method works by scoring time scales based on how well their resulting 2D embeddings result in maximally circular holes in the resulting attractor. This is achieved by applying a delay embedding on randomly sampled short trajectory strands, computing its persistent homology and scoring each identified hole (1-dimensional homology) according to its circularity and its efficient use of points.

We would like to thank all the reviewers again for their comprehensive feedback and taking the time to review our work. We believe that the changes made to the submission have improved the quality of the manuscript and hope that they have also sufficiently addressed the reviewers' concerns. Should there be any new or additional concerns, please do not hesitate to contact us for clarification.

Kind Regards,
Eugene Tan, Shannon Algar, Débora Corrêa, Thomas Stemler, Michael Small
The University of Western Australia
Mon 9$^{\text{th}}$ Oct, 2023

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I thank the authors very much for properly considering all my comments. The authors have further improved their manuscript with comments from other reviewers. I am happy to reiterate my positive impression of this work and warmly recommend its publication in the present form.

Reviewer #4 (Remarks to the Author):

The authors have revised the manuscript in a satisfactory manner. Thus, I recommend its publication after the following minor correction is made:

1. P.33, Eq. (D1), L_{ij} seems R_{ij}.