# An unexpected discrepancy in a well-known problem: Kraskov estimators applied to spiking neural networks

Pedro A.M. Mediano and Murray Shanahan

Department of Computing, Imperial College London. 180 Queen's Gate, London SW7 2RH

pmediano@imperial.ac.uk; m.shanahan@imperial.ac.uk

#### Abstract

While revisiting a problem in spiking neural networks (Shanahan, 2008) we discovered an interesting anomaly that we feel deserves further investigation as it could signify a useful contribution to the field of complex systems analysis. Robust methods for estimating information-theoretic measures like the two mutual information estimators in Kraskov et al. (2004) normally yield similar results, but we have found a case where they do not. Through extensive testing and the use of software libraries that are known to be reliable, we are confident that this seemingly contradictory outcome is not an error, and that it is a new discovery. We advocate the need for more thorough testing and benchmarking of estimators, and warn against the risks of drawing premature conclusions when using only one test method.

# Introduction

The increasing availability of data and computational power are driving research in complex systems, notably in pioneering interdisciplinary fields like Artificial Life and Computational Biology. Therefore the study of complex systems has become even more vital. Reputable studies strengthen their claims using statistical analyses yet there is an inherent risk when using only one test method. This was highlighted beautifully in an Ig-Nobel-winning article that deliberately measured brain activity in a dead salmon (Bennett et al., 2011).

This vulnerability makes using multiple tests more appealing; perhaps the more the better. Naturally if the results differ between methods, this needs explaining. We have found a discrepancy in the outcome of the two current best-ofbreed non-parametric estimators for information-theoretic measures introduced by Kraskov, Stögbauer and Grassberger (2004). This is interesting because the two estimators, known here as KSG 1 and 2, are almost equivalent — KSG 1 has higher bias, KSG 2 higher variance — and normally exhibit comparable behaviour (see Figure 1). Unexpectedly, the left-hand side of Figure 2 shows a peak in one line. This paper analyses and discusses this anomaly.

#### Methods

We discovered this anomaly when revisiting the problem of relating the dynamical complexity to the structural complexity of spiking neural networks (Shanahan, 2008). An intriguing property of biological neural networks (brains) is their impressively rapid and seemingly paradoxical ability to produce a single coherent response despite small groups of neurons appearing to act individually in parallel. We therefore want to measure the equivalent behaviour in our system by seeing how it balances local and global information (in the general sense of the word). Reassuringly, this property can be measured independently from two diverse perspectives: network connectivity (which components connect to others) and network activity (how much traffic they send and receive). For our purposes, network activity equates to dynamical complexity and network connectivity to structural complexity. Understanding our assessment of these characteristics is helped if we first describe our system.

The system is a modular network of spiking neurons consisting of 8 modules of 100 excitatory neurons each. Each module has strong intra-connectivity and the inter-module coupling is achieved by adding long-range connections, the quantity of which is controlled by the parameter p which represents the fraction of long-range synapses between neurons in different modules. These 800 neurons are connected to a common pool of 200 inhibitory neurons so that the 8 clusters compete in a winner-take-all fashion. The model and parameter values for all neurons are from Izhikevich (2003), a wellestablished standard in the field. To sustain activity in the network, we supply every neuron strong Poisson-distributed current pulses with a rate parameter of 10 Hz.

We assess structural complexity using the small-world index S (Humphries and Gurney, 2008), calculated for the binary directed network formed by the 800 excitatory neurons. S quantifies the trade-off between high local clustering and short path length. We calculate all the relevant structural complexity measures using the Brain Connectivity Toolbox by Rubinov and Sporns (2010).

Similarly, we assess dynamical complexity using neural complexity  $C_N$  (Tononi et al., 1998). Neural complexity captures the balance between segregation and integration in a system by considering how the elements interact at all possible scales, from the smallest component up to the entire

system. This interaction is measured using mutual information, the building block of  $C_N$ , that in our experiments is directly estimated with either KSG 1 or 2.

We calculate all relevant information-theoretical measures using the free open source software package JIDT (Lizier, 2014). JIDT has been thoroughly validated against four other open implementations of KSG 1 and 2 so far. We use Kraskov et al.'s (2004) recommended setting k = 4, which is the default for JIDT.

We run the simulation for 200 s with a resolution of 1 ms. No measurements are recorded for the first 5 s to eliminate transient effects. The activity of the network is then recorded for 200 trials, measuring the firing rate of each cluster over a moving 50 ms window sampled at 20 ms intervals.

### Results

We calculate S and  $C_N$  for every trial and show the relation between the two and the model parameter p in Figures 1–3. Figure 1 reveals a strong correlation between S and  $C_N$ ; therefore we conclude that dynamical complexity as measured by  $C_N$  is strongly correlated with structural complexity as measured by S. Figure 3 shows that S has a unique maximum at an intermediate value of the parameter p, indicating there is an optimal trade-off between high clustering and low path length. Figure 2 is more interesting to interpret and is the main focus of this paper.

In the mid- and high-p region, where the modular structure of the network is less pronounced, both KSG methods give similar results. This suggests that the neural complexity is smaller the more disordered the network.

In the low-*p* region, however, the methods yield qualitatively different results. The main difference between the two is that KSG 2 shows a peak in  $C_N$  at around p = 0.05, whereas KSG 1 displays monotonically decreasing behaviour.<sup>1</sup>

# Discussion

In this section we describe the validation process we have taken and discuss the reasons that lead us to believe this case represents a genuine discrepancy between the methods and not an error in our procedure. After extensive searching, we believe this to be a newly reported discovery.

It is known that the KSG estimators are slightly different, which becomes more apparent when dealing with smaller datasets: the smaller the dataset, the more you can expect a discrepancy, and the larger that discrepancy might be. Therefore it is vital to check that this difference is not merely a finite sample size effect. We test the effect of dimensionality and sample size by comparing both methods on synthetic data. We generate multivariate normal random samples of the same covariance, dimensionality and size as the



Figure 1: Neural complexity of the modular network calculated with both KSG estimators. When plotted against S, both methods show similar behaviour.



Figure 2: Neural complexity of the modular network calculated with both KSG estimators. When plotted against p, the methods show disagreement in the low-p region.



Figure 3: Small-world index of the modular network as a function of the fraction of long-range synapses *p*.

<sup>&</sup>lt;sup>1</sup>Note that p = 0 does not imply a fully disconnected network — the excitatory clusters still interact via the common inhibitory pool, despite having no direct excitation between them.

simulated data and apply both KSG 1 and 2. In this case, both methods show identical behaviour, with an average difference below 0.1%. Therefore, dimensionality and sample size are discarded as an explanation for the anomaly.

One plausible hypothesis is that the discrepancy is a consequence of some unknown bias that affects KSG 1 and 2 differently. To discard this possibility, we use several surrogate data testing methods to correct for different biases. Nevertheless the difference remains after applying any surrogate data method in our suite (Lucio et al., 2012). In fact, the results shown in Figures 1 and 2 are corrected using amplitude-adjusted Fourier transform. We conclude that this difference is more fundamental; or at least is not one that conventional surrogate data testing methods can correct for.

Another (unlikely) hypothesis is that these particular sampling choices are somehow pathological, and a different way of collecting the data would solve the problem. To test this, we apply different low- and highpass filters on the raw spike time series by changing the moving window size and step and by taking up to two time differences of the averaged data. Again, these procedures preserve the difference between the outcomes of both methods.

Finally, to verify that this is not an artifact of the factors that make up  $C_N$ , we use the same procedure to estimate other information-theoretic measures. For example, the same discrepancy still holds if we estimate simpler quantities — like the mutual information between any bipartition of the network — or other more involved quantities — like interaction complexity  $C_I$ , an approximation of  $C_N$  also introduced in Tononi et al. (1998).

# Conclusions

KSG estimators, although generally accepted and broadly used in the analysis of complex systems, can yield inconsistent results when applied to more realistic cases that deviate from commonly used benchmarks.

The estimators produce qualitatively different results on our spiking neural network data, despite giving almost identical results when applied to an artificial dataset of the same size and dimensionality as the data extracted from the network. This difference is consistent and is obtained after both filtering the data and correcting the results using surrogate data methods.

This difference is important. An experimenter, Alice, using only KSG 2 to analyse her data will come to the conclusion that there is a peak of "interestingness" in a point where the network is not completely disconnected but still retains a strong modular structure; while her colleague Bob analysing the exact same data using only KSG 1 will conclude that the more segregated the network, the more complex behaviour is obtained. However, since both methods perform equally accurately in a common benchmark, there is no good reason to believe one of the two researchers is right while the other is wrong. The benchmarks commonly used when testing estimators are helpful and necessary, but alone they are not sufficient to guarantee the validity of an estimator in other contexts. Real data on which the estimators are applied are often much less well-behaved than the synthetic data used in benchmarks. It is crucial to develop rigorous methods for testing, understanding and comparing estimators.

# Acknowledgements

The authors gratefully acknowledge Brian Mitchell for his assistance with earlier versions of this document. This work is supported by an EPSRC doctoral training grant.

# References

- Bennett, C., Baird, A., Miller, M., and Wolford, G. (2011). Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction. *Journal of Serendipitous and Unexpected Results*, 1:1 – 5.
- Humphries, M. D. and Gurney, K. (2008). Network 'small-worldness:' a quantitative method for determining canonical network equivalence. *PLoS ONE*, 3(4):e0002051.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council, 14(6):1569–72.
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, 69(6):066138.
- Lizier, J. T. (2014). JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems. *Frontiers in Robotics and AI*, 1:37.
- Lucio, J. H., Valdés, R., and Rodríguez, L. R. (2012). Improvements to surrogate data methods for nonstationary time series. *Physical Review E*, 85(5):056202.
- Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3):1059–69.
- Shanahan, M. (2008). Dynamical complexity in small-world networks of spiking neurons. *Physical Review E*, 78(4):041924.
- Tononi, G., Edelman, G. M., and Sporns, O. (1998). Complexity and coherency: integrating information in the brain. *Trends* in Cognitive Sciences, 2(12):474–484.