

# A CORRELATION INTEGRAL BASED METHOD FOR PATTERN RECOGNITION IN SERIES OF INTERSPIKE INTERVALS

M. Christen, A. Kern and R. Stoop

Institute of Neuroinformatics,  
University / ETH Zürich,  
Winterthurerstrasse 190, 8052 Zürich, Switzerland  
e-mail: ruedi@ini.phys.ethz.ch

**Abstract**— *Methods for detecting patterns in interspike interval (ISI) series, obtained from biological neural networks, are usually template based. This approach faces the major problem, that a priori assumptions about possible pattern structures have to be made. To avoid this difficulty, we propose a novel statistical approach, based on the correlation integral. Applications to model and neuronal data show the reliability of the method, even in noisy conditions.*

## I. INTRODUCTION

Biology can be viewed as an alternative information processing paradigm, that often proves far more efficient than conventional approaches. Whereas the underlying processes (neurons and their connections by means of synapses) can accurately be modeled by electronic circuits [1], their basic way of functioning is not yet understood. However, it seems evident that these circuits work on distributed parallel processing principles, and that information is encoded differently than it would be in traditional signal processing. In biological information processing systems, generated activity manifests in neuronal firing events, called spikes. Temporal recording of firing events provides ISI series as the empirical material to work on. It can be expected that the information processed in the network, is encoded by some structure in the ISI series. This implies pattern recognition as a first fundamental step in our investigation of biological information processing. One of the most obvious structures to look for, is patterns in single ISI series. Generally, ISI patterns can be defined as parts of the series that appear significantly more often than in random distributions [2]. For the identification of patterns, usually a set of template patterns is predefined and their occurrence within the ISI series is counted (*template-based methods*, [3]). In neuronal data, such patterns composed of one, two, three, four and five consecutive ISI have been reported [4].

However, template-based methods suffer from two

fundamental difficulties: First, the detection relies on the set of pre-chosen pattern templates. As the patterns are *a priori* unknown, the inclusion of the appropriate template is a matter of luck. Unbiased guessing will therefore require exceedingly large template sets. The second problem is the omnipresence of noise in natural processes. In biology, patterns cannot be expected to repeat perfectly, which leads to the problem how to choose the accuracy required for template matching. To avoid these difficulties, we propose a purely statistical approach for pattern detection. This approach is based on the correlation integral.

## II. PROPERTIES OF THE METHOD

### A. The correlation integral

Originally, the correlation integral was designed for the determination of correlation dimensions [5]. This application will not be considered in this report. Rather, we will explore the potential of correlation integrals for the detection of ISI-patterns.

Consider an arbitrary scalar time series of measurements  $\{x_{t_k}\}$ . From the data, embedded points  $\xi_k^{(m)}$  are generated using the coordinate-delay construction  $\xi_k^{(m)} = \{x_{t_k}, x_{t_k+\tau}, \dots, x_{t_k+(m-1)\tau}\}$ , where  $m$  is called the embedding dimension and  $\tau$  is a delay [6]. In the case of interspike interval data, however, no explicit delay is involved. By setting  $\tau = 1$ , the construction can nevertheless be performed [7]. From the embedded data, the *correlation integral* is calculated

$$C_N^{(m)}(\varepsilon) = \frac{1}{N(N-1)} \sum_{i \neq j} \theta(\varepsilon - \|\xi_i^{(m)} - \xi_j^{(m)}\|),$$

where  $\theta(x)$  is the Heavyside function ( $\theta(x) = 0$  for  $x \leq 0$ , and  $\theta(x) = 1$  for  $x > 0$ ) and  $N$  is the number of embedded points. The correlation integral  $C_N^{(m)}(\varepsilon)$  averages the probability of measuring a distance smaller than  $\varepsilon$  between two randomly chosen points  $\xi_i^{(m)}$  and  $\xi_j^{(m)}$ . In practical applications,  $\log C_N^{(m)}(\varepsilon)$  is often plotted against  $\log \varepsilon$  (the

so-called *log-log plot*). Since we will investigate the behavior of  $C_N^{(m)}(\varepsilon)$  for  $\varepsilon$  finite, there are no stringent limits in terms of data size.

### B. Emergence of steps in the log-log plot

Patterns will manifest as a clustering of the embedded data. For the calculation of  $C_N^{(m)}(\varepsilon)$ , an embedded point  $\xi_0^{(m)}$  is randomly chosen. Then, the number of points in its neighborhood  $U(\xi_0^{(m)}, \varepsilon)$  is measured, as  $\varepsilon$  is enlarged. If the point belongs to a cluster, upon an enlargement of  $\varepsilon$  many points will join the neighborhood. I.e.,  $C_N^{(m)}(\varepsilon)$  quickly increases. Once the cluster size is reached, less points will be recruited, which leads to a locally flat part of the log-log plot. Proceeding in this way, step-like structures emerge. The denser the clustered region, the more prominent the step-wise structure. Noise tends to smear the underlying patterns and, as a consequence, the step-like structures. A diminished inclination of the steep parts of the steps can be taken as an indication of additive noise. A non-zero slope of the locally flat parts indicates instability of the pattern within the series.

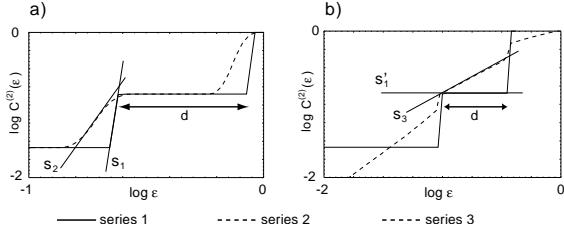


Fig. 1. Log-log steps, pattern precision and stability of occurrence. Steeper steps indicate more precise patterns ( $s_1$  vs.  $s_2$  in case a). More horizontal levels emerge for more stable patterns ( $s'_1$  vs.  $s_3$  in case b). In both cases,  $d \sim 0.58$  and  $m = 2$ .

To corroborate this by examples, we constructed three ISI series from the ISI set  $\{1,2,4\}$ , where the given numbers can be taken as ISIs measured in ms. The first series were generated as a repetition of the sequence  $\{1,2,4\}$ . To this series, indexwise noise of maximal  $\pm 1\%$  of the smallest ISI was added. To obtain the second series, the noise was increased to  $\pm 10\%$ . The third series was obtained by appending either the sequence  $\{1,2,4\}$  (with probability  $p = 0.5$ ), or a sequence consisting of three intervals randomly drawn from  $I = (0, 4)$ . For all obtained series, the correlation integrals were evaluated at fixed embedding dimension  $m = 2$  (the influence of the embedding dimension will be discussed below). Throughout the investigation, we used the maximum norm as the distance measure. This not only speeds up the com-

putation, it also scales well for the comparison among different embedding dimensions [8]. Degeneracies introduced by this choice are removed upon the addition of a small amount of noise. If we compare the results of the first with the second series, we observe that the “noisier” pattern exhibits a smaller slope  $s_2$  vs.  $s_1$  in Fig. 1 a). Comparing the results from the first vs. the third series, we observe a smaller slope  $s'_1$  for the more stable pattern vs.  $s_3$ , see Fig. 1 b). Both comparisons corroborate our predictions [9].

In practical applications it is generally difficult to clearly distinguish between the two aspects, as they naturally interfere. And often, if steps are smeared by noise, the derivative of the log-log plot is a clearer indicator for the presence of patterns. In the corresponding derivative plot, narrow peaks separated by intervals where the derivative  $\Delta \log C_N^{(m)}(\varepsilon)$  is close to zero, indicate precise and frequent replicating patterns. The number of peaks then corresponds to the number of steps. To improve the indicator, it has proven useful to consider the sum  $\sum_{i=1}^m \Delta \log C_N^{(i)}(\varepsilon)$  (see [8]).

Our observations are also valid in less simplistic settings. To show this, we extended our investigations to data composed of the ISI sequences  $\{10, 1, 100, 1, 100, 5, 2, 1\}$ ,  $\{2, 6, 10, 10, 20\}$ ,  $\{5, 6, 6, 2, 6, 5\}$ ,  $\{4, 2, 400, 1\}$ ,  $\{8, 2, 1, 5, 3, 3\}$ ,  $\{2, 7, 5, 700, 5, 2\}$ . ISI series were generated from these sequences according to the following procedures: a) Random selection of intervals from joined sequence, with additive noise uniformly selected from the interval  $[0, 2]$ . b) Random selection of a sequence, where the intervals were modulated by sinusoidal driving, and noise of same magnitude as in case a) was added. c) Random selection of a sequence without noise. In this way, the ISI series had identical probability distributions, whereas only in the series b) and c), patterns were present. The results shown in Fig. 2 confirm that steps, respectively peaks, only appear in the plots where patterns are present.

### C. Embedding dimension dependence

It is important to note, that from the number of steps or peaks present in the log-log plot, exactly what patterns are contained in the ISI series cannot be deduced. However, an estimation of the pattern length can be given according to the following arguments. For data generated by repeating a sequence of length  $n$ , the number of steps in the log-log plot,  $s(m, n)$ , can be expected to decrease with increased embedding dimension  $m$ . This prediction can be motivated

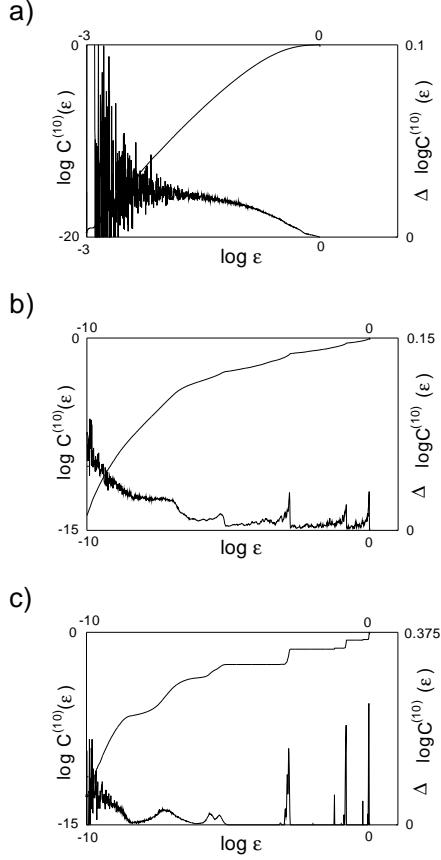


Fig. 2. Patterns manifesting as step-like behavior in the log-log plot (y-axis:  $\log C^{(10)}(\varepsilon)$ ) and as peaks in the derivative plot (y-axis:  $\Delta \log C^{(10)}(\varepsilon)$ ), respectively. Thick line: log-log plot; thin line: derivative plot,  $m = 10$ . a) Random selection of single intervals from the joined sequence: no steps visible. b), c) Steps emerge, if whole patterns are randomly chosen (see text).

as follows: When calculating the distance between two points, corresponding coordinates form a set of pairs, that can be ordered according to the absolute size of their differences. According to the maximum norm, the distance between the points is the largest difference found. As an increase of the embedding dimension yields ever more pairs, thus the presence of a particularly large difference (which has an increased probability for being present in higher dimensions) will dominate the distance and suppress the occurrence of smaller distances. The analytical formula of  $s(m, n)$  is beyond the scope of this paper. The number of steps can, however, be computed numerically, if all differences between the  $n$  numbers forming the pattern are distinct (see Table 1). The numbers obtained, clearly corroborate the expected decrease of the number of steps as a function of the embedding dimension  $m$ .

As a tractable example, we investigated the ISI se-

	Pattern size $n$								
	1	2	3	4	5	6	7	8	
Embedding dimension $m$	1	0	1	3	6	10	15	21	28
	2	0	1	2	4	8	12	16	22
	3	0	1	1	3	6	9	12	17
	4	0	1	1	2	4	7	9	13
	5	0	1	1	2	2	5	7	10
	6	0	1	1	2	2	3	6	8
	7	0	1	1	2	2	3	3	7
	8	0	1	1	2	2	3	3	4
	9	0	1	1	2	2	3	3	4

TABLE I  
Maximum number of steps  $s(m, n)$  as a function of the embedding dimension  $m$  and pattern size  $n$ .

ries generated from replicating the sequence  $\{0.05, 0.24, 0.37, 0.44, 0.59\}$ . In embedding dimension 1, all ten possible differences are detected (see Fig. 3). Upon an increase of the embedding dimension towards 5, the number of steps detected decreases. After reaching the value 2, the number of steps remains constant, as predicted by Table 1.

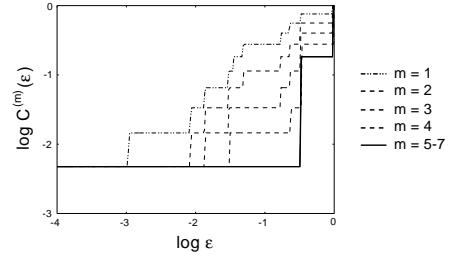


Fig. 3. Decrease of the number of steps as a function of the embedding dimension  $m = 1, \dots, 7$  (sequence length: 5). At  $m = 1$ , 10 steps are discernible. This number decreases in full agreement with Table 1. Noise of maximal  $\pm 1\%$  of the smallest interval was added to the data.

Real data is always affected by noise. In this case, the effects are less clear-cut, and Table 1 provides a guideline only. Moreover, another basic difficulty complicates the determination of the pattern length. E.g., its mechanism can be illustrated as follows: If one step is present in the log-log plot, this can be due to either one pattern of 2 ISIs, or two “patterns”, of one ISI each. A greater number of steps complicates this problem further. One can, however, expect that the most pronounced step appears when the embedding dimension equals the pattern size. This is, because at this dimension, the pattern appears in its most complete form (neither cut into pieces, nor spoilt by

points that do not belong to the pattern). As a consequence, evaluating the log-log plot across different embedding dimensions provides a helpful indicator for the pattern length (details omitted).

### III. APPLICATION OF THE METHOD

We applied the method to ISI series from the lateral geniculate nucleus (LGN) of anesthetized cats [10]. We analyzed the responses of two cells to three classes of stimuli: 1) sinusoidal gratings with randomly changing spatial frequency and direction; 2) cartoon movies; 3) natural stimuli videos. Each class contained 2-4 stimuli of 10-30 seconds duration each. 15-25 repetitions of stimuli from one class were applied randomly. From these recordings, we composed six, cell and stimulus specific, ISI series. To avoid artificial patterns due to stimulus onset, the first 50ms of each trial were deleted. For stimulus 3, this led to an ISI series of length 2000 for neuron *a*, to 3400 for neuron *b*, and to lengths between 6000 and 12000 for stimuli 1 and 2. For the analysis, we used the derivative plot, where we summed over the embedding dimensions 1 to 7 and verified the independence of the results on the particular composition of the ISI series.

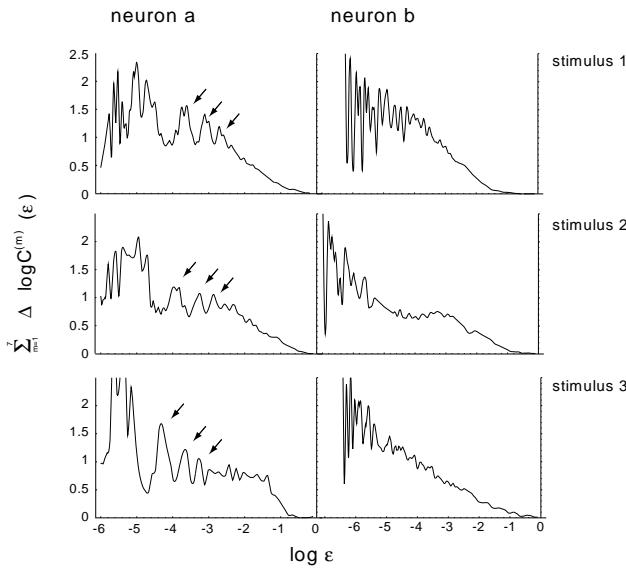


Fig. 4. Cat LGN data. Neuron *a*: Patterns emerging for all three stimuli (three characteristic peaks with preserved spacing between the peaks). Neuron *b*: Absence of patterns, for all stimuli used. In the plots, we display the cumulated derivative  $\sum_{m=1}^7 \Delta \log C_N^{(m)}(\varepsilon)$ .

The emergence of three characteristic peaks for all three stimuli shows the presence of patterns in the ISIs generated by neuron *a* (see Fig. 4, where different correlation integral normalizations are responsible for the shift of the peaks). For neuron *b*, the absence of peaks

indicates its inability to produce patterns. This is in agreement with earlier observations (see [11]) proposing a stimulus-independent categorization of neurons according to their ability / inability to fire in patterns.

### IV. CONCLUSION

The properties of correlation-integral pattern recognition can be summarized as follows: 1) The approach allows the unbiased testing for patterns, where only a small data set is required (for our investigations, 2000-5000 points were found to be sufficient). In its simplest application, our method may therefore be used for checking the existence of patterns. 2) Although the method does not directly deliver the patterns themselves, indicators for their sizes are provided. When the existence of patterns is indicated, this additional information (e.g., possible pattern size combined with the locations of the steps) can be used in combination with template-based methods, to reveal the patterns.

The authors thank Valerio Mante and Matteo Carandini for providing the LGN cat data used in this analysis. This work was supported by the SNF and by a KTI contract with Phonak AG.

### REFERENCES

- [1] C. Koch. *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, Oxford, 1999.
- [2] F. Rieke, D. Warland, R. de Ruyter van Steveninck, W. Bialek. *Spikes. Exploring the neural code*. MIT Press, Cambridge, London, 1999.
- [3] J. E. Dayhoff, G. L. Gerstein. Favored patterns in spike trains. I. Detection. In *J. Neurophysiol.*, 49(6), 1334-1348, 1983.
- [4] R. Lestienne, B. L. Strehler. Time structure and stimulus dependence of precisely replicating patterns present in monkey cortical neuronal spike trains. In *Brain Res.*, 437, pp 214-238, 1987.
- [5] P. Grassberger, I. Procaccia. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. In *Physica D*, 13, pp. 34-54, 1984.
- [6] J. Peinke, J. Parisi, O. E. Roessler, R. Stoop. *Encounter with Chaos*. Springer, Berlin, 1992.
- [7] T. Sauer. Reconstruction of dynamical systems from interspike intervals. In *Phys. Rev. Lett.*, 72, pp 3811-3814, 1994.
- [8] Our numerical algorithm rescales the distances between points by the largest existing distance, which leads to embedding dimension-independent results.
- [9] The distance between the steps is in both cases  $\sim 0.58 = \log_2 3 - \log_2 2$ , corresponding to the difference of the logarithmic distances between the embedded points  $\{1,2\}$ ,  $\{2,4\}$ ,  $\{4,1\}$ .
- [10] T. C. B. Freeman, S. Durand, D. C. Kiper, M. Carandini. Suppression without inhibition in visual cortex. In *Neuron*, 35, pp 759-771, 2002.
- [11] R. Stoop, D. A. Blank, A. Kern, J.-J. van der Vyver, M. Christen, S. Lecchini, C. Wagner. Collective bursting in layer IV – Synchronization by small thalamic inputs and recurrent connections. In *Cog. Brain Res.* 13, pp 293-304, 2002.