On Measurement of Internal Variables of Complex Self-Organized Systems and Their Relation to Multifractal Spectra

Dalibor Štys, Petr Jizba, Štěpán Papáček, Tomáš Náhlík, Petr Císař

On measurement of internal variables of complex self-organized systems and their relation to multifractal spectra

Dalibor Štys[1], Petr Jizba[2], Štěpán Papáček[1], Tomáš Náhlík[1], and Petr Císař[1]

[1] School of Complex Sytems, Faculty of Fishery and Water Protection, University of South Bohemia, Zámek 136, 373 33 Nové Hrady, Czech Republic
{stys,papacek,nahlik,cisar}@jcu.cz
[2] FNSPE, Czech Technical University in Prague, Břehová 7, 115 19 Prague, Czech Republic
p.jizba@fjfi.cvut.cz

Abstract. We propose a method for characterizing structured, experimentally observable, complex self-organized systems. The method in question is based on the observation that number of self-organized systems can be mathematically perceived as consisting of several interconnected multifractal components. We illustrate our key results with ensuing applications. The relation of the results obtained to known examples of strange attractors is also discussed.

Keywords: Rényi entropy, principal component analysis, generalized dimensions, multifractal spectra

1 Introduction

Self-organized (SO) systems typically refer to physical and biological systems in which patterns or structures present at the macroscopic level arise solely from a coherent interaction among lower level components of the system. The best known examples are undoubtedly living organisms which range in their complexity from simple cells through herds, flocks, insect colonies to humans and their herding behaviour or cities. There are many models used to generate analogies of specific features of such observable structures. The comparison of these models to observed dynamical systems is the question which awaits yet the answer. It is purpose of this paper to point out that some signatures that are met in analysis of SO systems may be inferred by the statistical approach based on a multivariate analysis. This approach has at the moment a widely utilized version epitomized by the principal component analysis (PCA) alongside with many new developments including, e.g., analyses developed for data exhibiting distribution skewness [1]. It should be, however born in mind that many realistic intensity distributions found experimentally in images of SO systems such as Belousov–Zhabotinsky reaction (see Fig. 1) or living cells (see Fig. 2) are rather complex

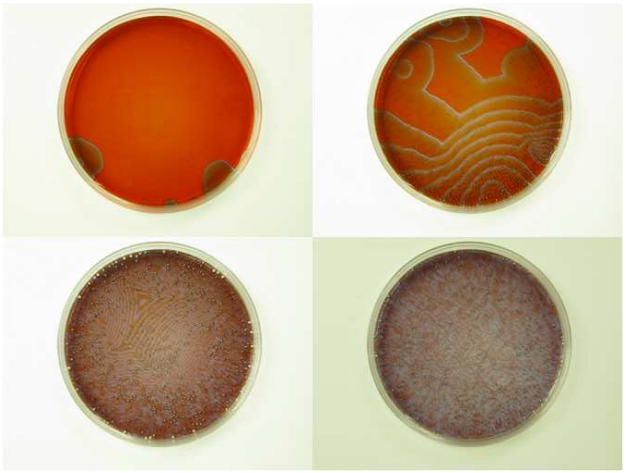

Fig. 1. Examples of structures observed in the course of Belousov–Zhabotinsky reaction. From the upper right to bottom left we depict images of 38th, 140th, 277th and 425th minute of the reaction trend.

and cannot be mapped on a simple (skewed) normal distribution. Thus, the question of multivariate analysis on data-sets coming from SO systems remains open, regardless of its wide utility, for instance, in various fields of biology and medicine [2].

In this paper we describe a surprisingly successful application of the PCA for charting the state space of the Belousov–Zhabotinsky reaction. The key message which we wish to convey here it the *modus operandi* which illustrates the basic logical steps that should be utilized on the way to the state-space description of self-organized systems. In passing, we will also see what potential obstacles can be expected along the route.

2 Measurement in self organized systems

2.1 The problem of orthogonal coordinates

The principal component analysis has been first introduced by Pearson as early as in 1901 [4]. Its application consists of three steps: (1) normalization of each data-set, (2) calculation of the covariance matrix and (3) calculation of principal components, i.e. orthogonal coordinates which determine directions of main variance in the data. When principal components are calculated, it may be calculated covariance of each of the original data-sets with each principal components. These covariances are called *loadings*.

Experimentalist's experience is often such, that loadings are naturally coinciding with components of one known process or feature. For example, chemical compounds characterizing one biochemical process have highest loading on the

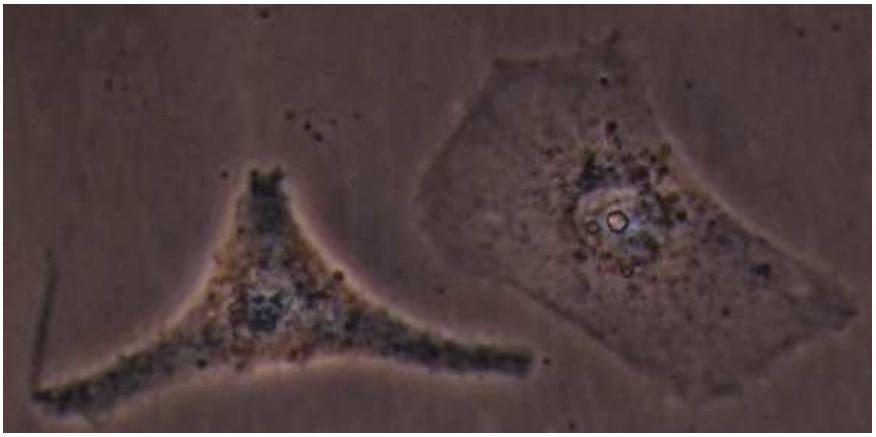

Fig. 2. Two typical structures of living HeLa cell observed by the microscope. The figure comes from Refs. [12, 13].

same principle axis. In Ref. [12, 13] we have discussed the necessity to characterize microscopic images of living cells in chemical, as well as, in physical space. The chemical space needs to be characterized with chemical potentials which are in logarithmic relation to activities. In equilibrium thermodynamics, both these variables are related to standard thermodynamical variables of temperature, pressure and volume through systems state function.

It is somehow naturally assumed, that pressure, volume, temperature and number of molecules are orthogonal state coordinates of the system. At least, that is the impression from textbooks on physical chemistry. In fact, the only mathematically correct statement is that state function of the ideal gas is a manifold in the orthogonal state space with coordinates pressure, volume, temperature and number of particles. For our discussion it is, however, more valid to discuss the Gibbs free energy. For ideal gas, the state function is a plane — eventually hyperplane, in the space with coordinates G, T and $\wp_i$ (i.e., Gibbs energy, temperature and partial pressures of chemical components in the system). With some manipulations we for the condensed phase case come to widespread formula

$$G \;=\; G_0 \;+\; RT \sum_{i=1}^{n} \ln c_i \,, \tag{1}$$

where c_i are concentrations of system components. G_0 denotes standard Gibbs energy of given system and R the universal gas constant. All equilibrium thermodynamics is devised in this state space. In chemical thermodynamics, in order to obtain more realistic description, activity a_i (or fugacity) replaces the partial pressure of ideal gas component. The correction factor, so-called activity coefficient γ_i, is in function accounting for transformation of coordinates in order to

obtain formally the same equation as that for ideal gas.

$$G \;=\; G_0 \;+\; RT\sum_{i=1}^{n}\ln a_i \;=\; G_0 \;+\; RT\sum_{i=1}^{n}\ln c_i + RT\sum_{i=1}^{n}\ln\gamma_i\,, \qquad (2)$$

Activity coefficient is non-linear function of concentration of all species present in the sample. From physico-chemical point of view, it accounts for all molecular interactions in given system. Generally, manifolds of state functions are accessible only experimentally even for such simple systems as two component mixtures of miscible liquids and their vapours.

Principle components approach [4] looks at the problem from different point of view. The resulting variables, principal components, may be assumed merely as practical descriptors of the system. When the problem is treated from more elementary point of view, the principal components are best approximation to internal orthogonal coordinates in state space appropriate to stochastic (macroscopic) variables of which each has normal distribution. Instead of constructing the state function as manifold in a given coordinate system, we find coordinate system in which the state function is a plane in multidimensional space. When equilibrium physico-chemical system is examined, we may consider principal components best approximations to log-activities as functions of all systems variables.

2.2 Orthogonal coordinates of in the phase space and generalized fractal dimensions

A nice and extensively studied example of self-organizing object is the Lorenz attractor (cf., e.f. Ref. [7]). The beautiful butterfly wings structure is plotted in orthogonal phase-space coordinates which originally had physical meaning of total flux of matter, temperature difference between incoming and outgoing stream and deviation of temperature gradient from linearity. Certainly, these are not parameters which are easily measurable in model Bénard cell, where is to the measurement accessible the diameter (in general shape) of the cell and temperature of the lower and upper plate.

Results of measurements of real objects are to a large degree stochastic. The real system is subjected to external fluctuations and repeated use of for example - mechanical measurement device naturally brings up normal distribution of measured value probability. Digitization of analogue signal has similar effect, however, often leading to distortion of the signal. These are just few of many sources of stochasticity in real measurements.

The quest for proper measurement of self-organizing systems is to describe them by measurable variables which reflect their internal structures. As well as principal components best reflect the axes according to which a given dataset may be best separated. Self-organizing systems, however, are not homogeneous, they have their observable, self-maintained, dynamic structures. Thus the first problem to be addressed, before the multivariate analysis is performed, is to find a method how to characterize the internal structure.

We essentially measure what may be measured by accessible equipment. Or, in more precise expression, the measured system has to be examined only in relation to its phenomenological variables which are the measured sub-set of all possible variables representing system attributes, as was precisely formulated for example by Žampa [8] (see also [12]). The internal, in real system at least within statistical error orthogonal, coordinates, are not accessible to the experimentalist. We proposed earlier [12] to extract the information content of the given data-set using Rényi entropy gain approach. Rényi entropy has a close connection to generalized dimension of the multifractal objects [9, 10]. We, at given image resolution yardstick, use Rényi entropies at different parameter, calculated simply from occurrences of different signals, as first attempt to determine measure in the state space.

At this place we discuss self-organizing multifractal object whose stability is based on their character of chaotic attractor. This allows us to build direct analogy with the Lorenz attractor. Each individual multifractal contribution to resulting spatially distributed signal, i.e. image, has its own spectrum of generalized dimensions. There should be as many components as there are dimensions of the phase space.

Each component of the real multifractal object should have its own generalized spectrum. In our practical approach this means that each of the Rényi entropy values at each alpha has its occurrence in the image. We determine the information contribution of each data point to the object by calculation of difference in Rényi entropy of the data-set between the data-set including and excluding the examined object. This is the point information gain (PIG) $PIG_{\alpha,x,y}$ introduced recently in Ref. [12]. For given data-set, i.e. image, the three-dimensional spectrum in coordinates Rényi entropy, the dimensionless α coefficient and occurrence of a given values in the set is a unique characteristic of the image. We are well aware of the fact that the relation of PIG to individual component of the generalized spectrum of individual component of the observed SO is only vague, but we do not have better technical way how to proceed. PIG for given point x,y at given alpha is calculated according to formula

$$\mathrm{PIG}_{\alpha,x,y} \;=\; \frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{n} p_{i,x,y}^{\alpha}\right) \;-\; \frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{n} p_{i}^{\alpha}\right). \tag{3}$$

where $p_{i,x,y}^{\alpha}$ and p_{i}^{α}) are probabilities of occurrence of given intensity in the image without and with the examined point. In the first approach, calculation of PIG from whole image, distribution of intensities within the object of study is not considered. Then occurrence and entropy represent the same feature. Inclusion of spatial information may be done by defining the reference frame, in our case the cross whose shanks intersect in examined point. We shall show below that the later method allows in the examined case better estimation of orthogonal principle components.

In the next step we calculate integral information as potential state variable. We calculate sum of all PIG values, the point information gain entropy PIE, and point information gain entropy density PIE/points, the sum of all PIG levels

at given alpha. Unfortunately it is likely that in non-symmetrical systems PIG, PIE and PIE/ points is dependent on choice of coordinates. This is given by the measurement device, for illustration we may consider that we examine the structure of Lorenz attractor while having only one arbitrary plane from which it is observed.

$$\text{PIE/ points}_{\alpha} = \sum_{i=1}^{n} \text{PIG}_{\alpha,i} \,. \tag{4}$$

In this article we, on measurable systems, experimentally examine following hypothesis:

1. The orthogonal coordinates obtained by principle component analysis are practically usefull approximation to the internal orthogonal coordinates of the asymptotically stable system which gives rise to observed self-assembling object.
2. The covariances (loadings) between each calculated PIE/points or PIE values and the given principal component are reasonable and practically usefull approximation to intensities of spectral lines of individual multifractal component, i.e. chaotic attractor - in observed set which gives rise to observed self-assembling object.

3 Results

3.1 Observables in general projections of a strange attractor

An obvious way to determine validity of our approach is to prepare simulated data and expose them to the aforementioned analysis. Recent calculation of fractal dimension of the Lorenz attractor was done by Viswanath [5]. He reported that calculated values are very sensitively dependent on a numerical precision of the implemented calculation. Addition of noise would add additional, and surprising, dimension to such demanding calculation. This we wish to demonstrate on an illustrative example.

In Fig. 3 we show snapshots of the evolution of a toy model of Lorenz attractor [11]. There we show results of simulation in the absence of noise. We further show results of the same simulation with on-negative Gaussian noise applied in all three orthogonal directions. This toy model, in which the numerical integration is performed by the Euler method with a relatively large integration step, shows impressive robustness of the Lorenz attractor system of equations. The influence of Gaussian noise is in a sense puzzling: instead of generating random behaviour of the system it leads to *de-coherence*. When noise is applied in y axis it leads to preference of one wing of the butterfly and when applied in the z axis the system collapses into one point of focus. We do not provide any proof of generality of this observation, however, this result seems to be in agreement with observations of Jalnine *et al.* [15] who predicted collapse of chaotic behaviour in the van der Pol oscillator at increasing noise. Similar results using different approach to noise generation were also reported by Zahri [16]. The model is

freely available [11] and the reader is encouraged to explore it. In any case, it illustrates that any simple assumption about emergence of normal distribution may fail in the case of systems based on strange attractors. Obviously, however, the attempt to obtain image of internal orthogonal structure in chaotic attractor including noise would be extremely sensitive to computational performance and we refrained to include it into our analysis. We limit ourselves to experimentally observed SO systems, where the Nature calculates the trajectory for us with best achievable precision. Since we do not have any information on probability density function of the processes contributing to observed phenomena in experimental self-organizing systems, it is necessary to provide some logically feasible distribution. On the basis of the central limit theorem, it is natural to expect that the normal distribution will be the best choice. Under this assumption we use principal component analysis. As a first approximation to the state-space variables we use the Rényi entropy values for individual alpha parameters.

3.2 Measurable variables in projections of general strange attractor

In Fig.5 we show the basic steps of our analysis. We first calculate information structure of a given image. For each point we calculate the PIG, the Rényi entropy difference at given α calculated from experimentally determined probability density function including and not including the examined point [11, 12]. At the moment we have implemented two methods of sampling of the image

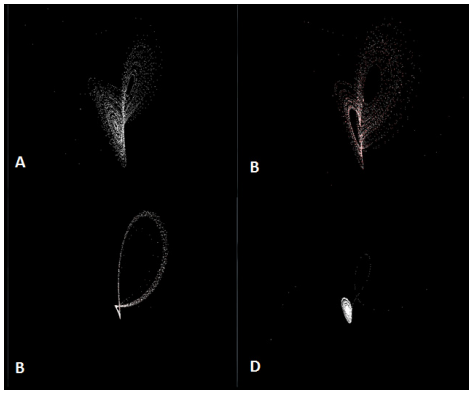

Fig. 3. Examples of behaviour of a toy model of Lorenz attractor. Model was prepared in Netlogo3D [17] program environment. Gaussian random noise was added in three dimensions separately. The noise was added as positive random number with normal distribution. Panel A shows original, noise-free, model. Addition of noise in x axis (panel B) shows expected randomisation of the trajectory. Addition of noise in the y axis, however, leads to formation of narrow stripe of trajectories (panel C) and addition of noise in the z-axis (panel D) leads to collapse of the chaotic attractor structure.

for the probability density function calculation. First is the obvious approach in which the number of points of given intensity is summed and normalised. In the second approach, the image is sampled for each of the points using a cross, whose shanks intersect in the examined point. The probability density function is calculated for the intensity distribution along the shanks. More detailed description of the approach was given in [13, 14]. We examined typical images in the course of self-organization of emergent structures of the Belousov–Zhabotinsky reaction.

The question remains whether PIE/points values may be used as coordinates in the phase space. This was examined experimentally. We utilized plain approach to computation of principal components as implemented in Unscrambler software [6]. One of successful attempts is depicted in Fig. 5, where we show decomposition of the system trajectory of the Belousov–Zhabotinsky into series of states which are asymptotically stable under current conditions. The reason why we can not provide one single recipe for the analysis originates comes from the process of observation itself. The geometrical setup affects the course of self-organization. From that follows that proper method of sampling of the image has to be adopted accordingly. In other SO systems, such as living cells or flocks, the effect of geometry may be less pronounces because the geometry is determined by objects themselves. There the example is, however, less instructive because the number of important principal components is typically six or more. In this respect we may only say that such a method was successfully tested on several typical data-sets but its generality has not been proven. In the last step we have examined the contribution of the original Rényi's fractal dimensions to the final ones. In Fig. 5 we depict the contribution of individual components to the whole image for 1st, 2nd and 3rd principal component. The first obvious thing is observation that there is not any qualitative difference between information carried in 1st and 2nd principal component. This may be seen from parallel course of dependency of individual PIE/points covariances when plotted against α. Some differences may be seen in 3rd principal component, where namely the contribution of the blue color deviates significantly. Following the line of arguments given in part in previous section we suggest that the principal components found for each asymptotically stable state is the best approximation to orthogonal coordinates of the chaotic attractor whose reflection are the observed measurable variables. By system state we understand, for example cluster found on the Belousov–Zhabotinsky reaction trajectory. Measurable variables in the case of self organizing systems must include also structure of the system. In the same way as in standard physico-chemical description where a function combining in principle all concentrations in the system makes up the true coordinate - activity, the determined PIE/points values combine into one principal component orthogonal inner coordinate describing macroscopic behavior of the self-organizing system.

Since we have started the analysis from the theoretical assumption of multifractality and calculated first technical approximation to Rényi's entropy, we may suggest that covariances between the PIE/points and principal components

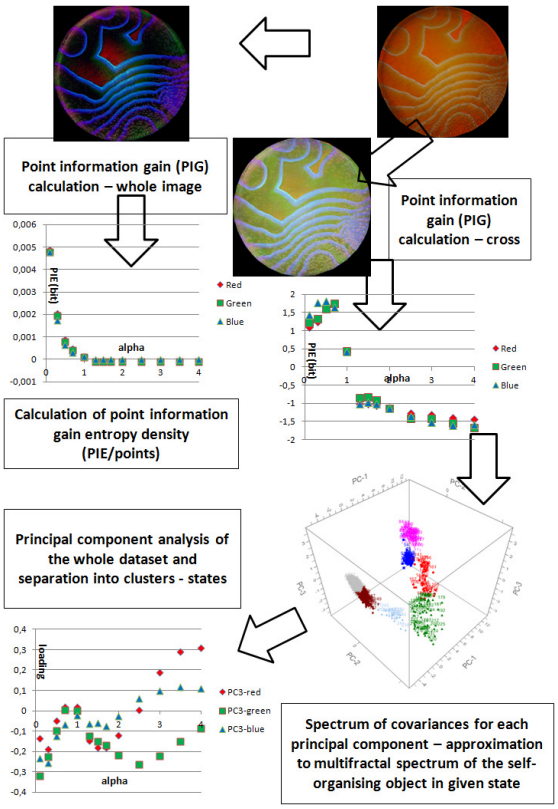

Fig. 4. Scheme of the analysis of experimentally observed self-organized systems. First, information on structure is extracted in the form of point information gain - PIG. Sum of all PIG levels provides point information gain entropy density - PIE/points. The vector of point information entropy densities is used as experimental dataset for the principal component analysis. Data clustering is performed on the data-sets which provides best experimentally achievable separation of the data-set into self-similar clusters - individual asymptotically stable states. For each of the clusters is performed separate analysis which provides, besides other information, spectrum of covariances between each principle component and original PIE/points value for given α parameter. These give, in our opinion, rough impression of how a generalized multifractal spectrum of the underlying attractor in projection to internal orthogonal component may look like.

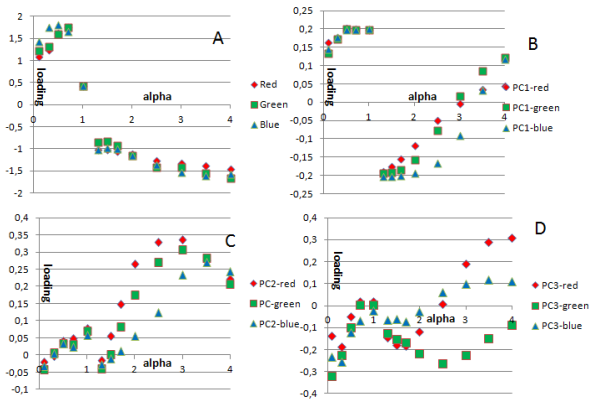

Fig. 5. Various approximations to generalized multifractal dimensions which arise in the course of the analysis. At panel A are given initial PIE/points values for the image 148 of the course of Belousov–Zhabotinsky reaction. Panels B, C and D provide covariances of first, second and third principal component and all utilized PIE/points values for the first cluster (images 56 - 135) which include the image under discussion. We propose that first, second and third principal components represent three independent contributions to the image.

are first technical approximation to multifractal spectrum of asymptotically stable dynamic systems determining systems dynamics. However, both above assumptions need further investigation. We provide all materials regarding the Belousov–Zhabotinsky reaction discussed in this article as well as all necessary software applications. Interested reader is encouraged to examine these primary data.

4 Materials and methods

4.1 BZ reaction

We used the kit provided by Dr. Jack Cohen http://drjackcohen.com/BZ01.html. The reaction mixture was composed out of following solutions: A. 25gm sodium bromate dissolved in 335ml water to dissolve to which 10ml of concentrated sulphuric acid (95-98%) was added, B. 10g sodium bromide dissolved in 100ml of water, 10g malonic acid dissolved in 100 ml of water, and D. 1, 10 phenanthroline ferrous complex (Fisons, Loughborough). The reactio was started in following order: 6ml of solution A was added to the glass beaker, then 0.5ml of B, and quickly mixed in 1ml of C. The brown mixture was left to lose bromine (by an open window) until it is pale straw colour or colourless (2-3 minutes if agitated or if in a flat dish). 1ml of the redox indicator D was added, mixed thoroughly and

poured into a 9cm glass or plastic petr dish, on a white (illuminated) background. The reaction was followed by colour camera Canon PowerShot G9.

4.2 HeLa cell microscopy

HeLa cells (Human Negroid cervix epitheloid carcinoma)were obtained from ECACC - European collection of cell cultures.The cells were grown at low optical density overnight in a 37 °C temperature in a synthetic dropout media with 30% raffinose as the sole carbon source. The nutrient solution for HeLa cells consists of: 86% EMEM, 10% newborn-calf serum, 1% antibiotics and antimycotics, 1% L-glutamine, 1% non essential amino acids, 1% NaHCO3 (all components from the PAA company).

In total, 5000 original phase-contrast images of growing cells were taken at 1-min intervals. The image collection was performed at 37 °C by using the Olympus IX 51 with an automated state, integrated incubator, and photographic Camedia C7070 camera. The objective with 20 magnification was used in an image capture.

4.3 Software packages

For calculation of PIG, PIE and PIE/points values were used software packages developed by the School of Complex Systems. Stable versions of these softwares are available at http://www.expertomica.eu/software.php and are constantly updated. Latest versions are available upon request. For calculation of the Lorenz attractor model was used modification of the model provided to the user community by Salzano in 2005 largely modified by Dalibor Stys [11]. The model is also available at http://www.expertomica.eu/software.php The principal component analysis was performed using Unscrambler software provided by CAMO company [6].

5 Acknowledgements

This work was partly supported and co-financed by the South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses (CZ.1.05/2.1.00/01.0024)

References

1. http://homepage.tudelft.nl/19j49/Matlab Toolbox for Dimensionality Reduction.html
2. Martens, H.: The Informative Converse paradox: Windows into the unknown. Chemometrics and Intelligent Laboratory Systems, 124-138 (2011)
3. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis. (Prentice Hall, London, 2007).
4. Pearson, K.: On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2, 559-572 (1901).

5. Viswanath, D.: The fractal property of the Lorenz attractor. J. Atmos. Sci. 20, 130141 (1963).
6. http://www.camo.com/
7. Lorenz, E.N.: Deterministic nonperiodic flow. Physica D, 190, 115-128 (2004).
8. Zampa, P., Arnost, R.: 4th WSEAS conference, 2004
9. Grassberger, P., Procaccia, P.: Characterization of Strange Attractors. Phys. Rev. Lett. 50, 346-349 (1983).
10. Grassberger, P., Procaccia, P.: Measuring the strangeness of strange attractors. Physica D 9, 189-208 (1983).
11. Stys at http://www.expertomica.eu/software.php
12. Stys, D., Vanek, J., Nahlik, T., Urban, J., Cisar, P.: The cell monolayer trajectory from the system state point of view. Mol. BioSyst. 7, 2824-2833 (2011)
13. Stys, D., Urban, J., Vanek, J., Cisar, P.: Analysis of biological time-lapse microscopic experiment from the point of view of the information theory. Micron. 41, 478-483 (2010)
14. Urban, J., Vanek, J.,Stys, D.:Preprocessing of microscopy images via Shannon's entropy. Pattern Recognition and Information Processing, 283-187 (2009)
15. Jalnine et al, arXiv:0805.0118v1
16. Zahri, M.: Numerical Solutions of a Stochastic Lorenz Attractor. J. Num.Mat. Stoch., 2, 1-11 (2010).
17. Wilensky, M.: (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.