

# Entropy modeling of sustainable development of megacities

A N Tyrsin<sup>1</sup> and G G Gevorgyan<sup>2</sup>

Ural Federal University, 620002, Ekaterinburg, Russia

<sup>1</sup>E-mail: at2001@yandex.ru

<sup>2</sup>E-mail: garnik.ggg@gmail.com

**Abstract.** The entropy approach of modeling multidimensional stochastic systems is described. It is based on the system representation as a multidimensional random vector and on the use of its differential entropy as a mathematical model. The possibilities of using this entropy model are considered for problems of monitoring the state of complex systems, including megacities, regions and critical infrastructure. Examples of practical implementation of the model are presented for the study of the sustainable development of megacities and regional environmental protection systems.

## 1. Introduction

The city is a complex ecological-socio-economic system consisting of a huge number of interacting elements. Large industrial centers are characterized by the enormous diversity and complexity of factors, elements of infrastructures, and the connections between them that affect the quality of its population life [1]. It should be noted that the quality of population life is a joint indicator reflecting the state of environmental protection, ecology, economy, education, health, housing conditions, culture, etc. These characteristics make it vital using a system approach for the sustainable development of these megacities, ensuring the growth of the quality of the population life [2]. A system approach involves the integration of the entire urban community into a system consisting of interrelated elements of infrastructures (environmental protection, ecology, economics, education, health, culture, politics, etc.), functioning of which is aimed to achieve the global goal of improving the quality of life of everyone [3].

However, despite the generally understandable main goal, the development of universal formal indicators characterizing the sustainable development of megacities is a complex problem. The existing diversity of various indicators that characterize various urban infrastructures and industries does not allow us to characterize the state of a megacity as a system.

Along with multidimensionality, another feature of the city as a system is its stochastic behavior. The stochasticity of the system is comes out in its probabilistic (unpredictable) nature of behavior, depending on random factors that can cause instability of individual parameters of the system as a whole. Peculiarity of such systems is the presence of a set of elements that are interconnected in a complex, probabilistic way. In this kind of situations, a multidimensional stochastic system is often modeled as a random vector.

One of the perspectives of modeling complex stochastic systems is based on the application of entropy. It is known that entropy is a fundamental property of all systems with an ambiguous or probabilistic behavior [4]. The concept of entropy is rather flexible and it can be clearly interpreted in terms of that specific area, where it is applied. It is being widely used in modern science to describe



the structural organization and disorganization, the degree of destruction of the connections between the elements of the system [5–8], including city systems [9, 10]. Therefore, it seems that the entropy could act as a universal parameter and it is ideal for solving the considered problems of the behavior of complex stochastic systems.

However, despite the frequent use of this term, the use of entropy for the modeling of open systems, unlike thermodynamics, is not formalized enough and is mainly of a qualitative and particular nature, there are no sufficiently simple and adequate mathematical models to link entropy with actual characteristics of states of multidimensional stochastic systems.

In these works the common thing is the use of information entropy:

$H(S) = -\sum_{i=1}^M p_i \ln p_i$ , where  $p_1, \dots, p_M$  the probability of the system  $S$  taking a finite number of corresponding values. The use of information entropy for the modeling of multidimensional open stochastic systems faces several serious difficulties. Let's list them.

- 1) The calculation of the information entropy requires an estimate of the  $p_i$  probabilities of the elementary states of the system. This requires large samples to ensure sufficient accuracy of entropy calculation.
- 2) Often there are difficulties, both with an unambiguous identification of a fixed finite set of states in a complex system, and with the fact that some states may not be known in advance.
- 3) Information entropy is not designed for the case of multidimensional systems. Therefore, it is difficult to model the relationships between the system elements.
- 4) The change in the variance of the studied process is not considered.

Therefore, the existing adequate entropy models of real objects are designed to solve particular problems.

The aim of the article is to consider another approach that allows one to universally construct and interpret entropy models of such complex stochastic systems as megacities.

## 2. The entropy model of a multidimensional stochastic system

The proposed approach is based on using the differential entropy [11]

$$H(\mathbf{Y}) = -\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_Y(x_1, x_2, \dots, x_m) \ln p_Y(\mathbf{x}) dx_1 dx_2 \dots dx_m \quad (1)$$

of the random variable  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$  with the  $p_Y(x_1, \dots, x_m)$  with the joint probability density function (PDF) without knowing its distribution type.

Let's represent the stochastic system  $S$  as a random variable  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ . Each  $Y_i$  element of the vector  $\mathbf{Y}$  is a one-dimensional random variable which is characterizing the functioning of the particular element of the system under study.

The entropy (1), being a functional from the set of probability densities of random variable  $\mathbf{Y}$ , is a number. And therefore it cannot be counted as an adequate mathematical model of a multidimensional system. In [12] it is proved that if all the components of  $Y_i$  have variances of  $\sigma_{Y_i}^2$ , then the  $H(\mathbf{Y})$  differential entropy (the entropy, from now on) of the random vector  $\mathbf{Y}$  is equal to:

$$H(\mathbf{Y}) = H(\mathbf{Y})_V + H(\mathbf{Y})_R = \sum_{i=1}^m \ln \sigma_{Y_i} + \sum_{i=1}^m \kappa_i + \frac{1}{2} \sum_{k=2}^m \ln(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2), \quad (2)$$

where  $\kappa_i = H(Y_i/\sigma_{Y_i})$  is the entropy indicator of the distribution type of the random variable  $Y_i$ ,  $R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2$  the coefficients of determination of regression dependencies.

The first two terms  $H(\mathbf{Y})_V = \sum_{i=1}^m H(Y_i) = \sum_{i=1}^m \ln \sigma_{Y_i} + \sum_{i=1}^m \kappa_i$  are the entropy of a random vector with mutually independent components and are called the randomness entropy, the third one –  $H(\mathbf{Y})_R = \frac{1}{2} \sum_{k=2}^m \ln(1 - R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2)$  – self-organization entropy, which characterizes the tightness of the joint correlation between the components of  $Y_i$ .

According to (2), the  $H(\mathbf{Y})$  entropy possesses threealism. There are three reasons for its change: the change of the level of scattering of its components, the change of the distribution type of its components, and the change of the strength of correlations between its components.

The advantage of formula (2) is that the entropy modeling of multidimensional stochastic systems on its basis doesn't require any definition of the distribution type of the multidimensional stochastic variable  $\mathbf{Y}$ , which is practically unfeasible in real problems. Herewith, unlike the methods of multivariate statistical analysis, the formal rigor and correspondence of model (2) to real experimental data is not lost. This allows us to use formula (2) to model and study real multidimensional stochastic systems and processes on limited experimental data set.

For multidimensional normally distributed random variable  $\mathbf{Y}$  we have [12]:

$$H(\mathbf{Y})_V = \sum_{i=1}^m \ln \sigma_{Y_i} + m \ln \sqrt{2\pi e}, H(\mathbf{Y})_R = \frac{1}{2} \ln(|\mathbf{R}|), \quad (3)$$

where  $\mathbf{R}$  is the correlation matrix of the random variable  $\mathbf{Y}$ .

According to (2), the parameters of the entropy model are:

- $\sigma_{Y_i}$  standard deviations of  $Y_i$  components,
- entropy indicators  $\kappa_i$  of distributions,  $i = 1, 2, \dots, m$
- $R_{Y_k/Y_1 Y_2 \dots Y_{k-1}}^2$  coefficient of determination of regression dependencies of the components of random vector  $\mathbf{Y}$ ,  $k = 2, 3, \dots, m$ .

The diagnostic model should explain the changes that occur in the studied object during functioning, in dynamics. Let's now consider the entropy of random vector from this point (2).

Let the stochastic system be presented as a random vector  $\mathbf{Y}$ . Then, based on the model (1) it is possible to monitor the state of the stochastic system by analyzing the change in its entropy. This can be done as follows. Let's assume that two random vectors and describe the previous and current periods of functioning of the system. We assume that the dispersions of all the components of random vector are finite.

To diagnose the state of a multidimensional stochastic system we'll adhere to the following steps [12]:

- 1) system behavior definition (stable/unstable), search for time dependencies of system behavior, and critical values;
- 2) detecting changes in the nature of the system ("randomness" or "self-organization") in critical periods;
- 3) analysis based on the detected reason: which element of the system turned to be the cause of the change in its state;
- 4) drawing conclusions about the impact of changes on the system, considering the identified critical points and their causes.

**Case 1:** First consider the case when distributions of all relevant components of  $Y_i^{(1)}, Y_i^{(2)}$  ( $i = 1, 2, \dots, m$ ) are described by the same type of distribution with some location and scale parameters. This means that  $\forall i \kappa_i^{(1)} = \kappa_i^{(2)}$ . Then the difference of entropies  $\Delta H(\mathbf{Y}) = H(\mathbf{Y}^{(2)}) - H(\mathbf{Y}^{(1)})$  is defined as:

$$\Delta H(\mathbf{Y}) = \sum_{i=1}^m \ln \frac{\sigma_{Y_i}^{(2)}}{\sigma_{Y_i}^{(1)}} + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}.$$

Having pair of components  $Y_i^{(1)}, Y_i^{(2)}$  with the same types, instead of threealism we'll have a particular case of dualism, i.e. the entropy can only change either due to the change of variance of the components of random vector or due to the change of the tightness of the correlation between these components.

The contribution of any  $l$ -th component in the change of the entropy of randomness:

$$\Delta H(\mathbf{Y})_{V,l} = \ln \frac{\sigma_{Y_l^{(2)}}}{\sigma_{Y_l^{(1)}}}, l = 1, 2, \dots, m.$$

Since  $R_{Y_m/Y_1 Y_2 \dots Y_{m-1}}^2 \geq R_{Y_m/Y_1 Y_2 \dots Y_{m-2}}^2 \geq \dots \geq R_{Y_m/Y_1}^2$ , the ultimate coefficient of determination  $R_{Y_m/Y_1 Y_2 \dots Y_{m-1}}^2$  most accurately describes the dependence of the  $Y_m$  component from the rest ( $m - 1$ ) components. Therefore, it is expedient to assess the contribution of the  $l$ -th component in the change of the self-organization entropy through the ultimate values of coefficients of determination.

$$\Delta H(\mathbf{Y})_{R,l} = \frac{1}{2} \ln \frac{1 - R_{Y_l^{(2)}/Y_1^{(2)} \dots Y_{l-1}^{(2)} Y_{l+1}^{(2)} \dots Y_m^{(2)}}^2}{1 - R_{Y_l^{(1)}/Y_1^{(1)} \dots Y_{l-1}^{(1)} Y_{l+1}^{(1)} \dots Y_m^{(1)}}^2}, l = 1, 2, \dots, m.$$

The total contribution of the  $l$ -th component in the change of entropy of the random vector is defined as  $\Delta H(\mathbf{Y})_l = \Delta H(\mathbf{Y})_{V,l} + \Delta H(\mathbf{Y})_{R,l}$ .

**Case 2:** Consider the general case when at least one pair of components  $Y_i^{(1)}$ ,  $Y_i^{(2)}$  are not described by the same type of distributions. Then the difference of entropies  $\Delta H(\mathbf{Y}) = H(Y^{(2)}) - H(Y^{(1)})$  is defined as:

$$H(\mathbf{Y}) = \sum_{i=1}^m \ln \frac{\sigma_{Y_i^{(2)}}}{\sigma_{Y_i^{(1)}}} + \sum_{i=1}^m (\kappa_i^{(2)} - \kappa_i^{(1)}) + \frac{1}{2} \sum_{k=2}^m \ln \frac{1 - R_{Y_k^{(2)}/Y_1^{(2)} \dots Y_{k-1}^{(2)}}^2}{1 - R_{Y_k^{(1)}/Y_1^{(1)} \dots Y_{k-1}^{(1)}}^2}.$$

Since in this case  $\kappa_i^{(2)} - \kappa_i^{(1)} \neq 0$ , there is now third factor of the entropy change caused by the distribution type change.

Thus, the case of the conservation of the distribution types of the random vector components is easier to implement and does not require the definition of the entropy indicators of the components. But the violation of this condition can cause significant errors in the assessment of the entropy dynamics, and, consequently, a decrease in the reliability of the diagnosis of the system state.

By monitoring the change of the  $\Delta H(\mathbf{Y})$  entropy as a whole and of its components, we can conclude about the state of the studied stochastic system and detect emerging trends in its changes. Analysis of changes in each component of the random vector  $\mathbf{Y}$  will reveal those of them having the greatest impact on change in entropy, and hence on change in the state of multidimensional stochastic system.

Since the city is an open system, its entropy can both increase and decrease. Systems with different randomness  $H(\mathbf{Y})_V$  and self-organization  $H(\mathbf{Y})_R$  entropies can have the same entropy values  $H(\mathbf{Y})$ . Therefore, it is necessary to consider the entropy (2) in vector form as:

$$h(\mathbf{Y}) = (h_V; h_R) = (H(\mathbf{Y})_V; H(\mathbf{Y})_R). \quad (4)$$

### 3. Practical use of the entropy approach in modeling the sustainable development of megacities

The sustainable development of a megacity as a complex system according to the proposed entropy approach consists in the simultaneous growth of diversity, opportunities for all elements of this system and the existence of a close relationship between these elements. This comes out with the fact that with the development of the megacity, its randomness entropy should gradually increase, and the self-organization entropy should decrease.

On the example of the analysis of the main indicators characterizing the state of critical infrastructures (environment, ecology, economy, education, healthcare, culture, etc.), we will consider the possibilities of practical use of the vector form (4) of the entropic model for studying the stability of the development of megacities and regions.

### 3.1. Comparative entropy analysis for the dynamics of Moscow, St. Petersburg and Ekaterinburg development in 1992–2015

The analysis will be carried out on the statistical dataset of Federal State Statistics Service [13]. From the large set of basic socio-economic indicators of cities, a system of 12 features is formed, characterizing all the main aspects of the city infrastructure:

- 1) The natural increase in population per 1000 people;
- 2) Employment-to-population ratio (in organizations), %;
- 3) Accrued nominal monthly average wages, thousand rubles;
- 4) The proportion of pensioners listed by social protection bodies, %;
- 5) The total area of residential premises, an average per an urban resident (at the end of the year), m<sup>2</sup>;
- 6) Number of pupils in pre-school educational organizations, thousand people;
- 7) Number of doctors per 1000 people;
- 8) The number of registered crimes per 1000 people;
- 9) Volume of industrial production, thousand rubles per capita;
- 10) The amount of work performed under construction contracts, thousand rubles per capita;
- 11) Retail trade turnover, thousand rubles per capita;
- 12) Investments in fixed assets, million rubles.

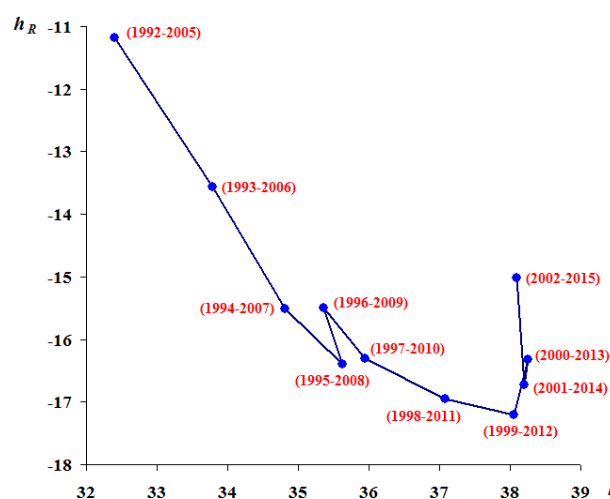
This system of indicators was formed according to the following rules. Pair correlations were considered between all the available in [13] main socio-economic indicators of cities, which are about twenty. It turned out that a number of indicators are very strongly correlated (coefficients of pair correlation exceeded 0.9). One indicator of these mutually correlated indicator groups was kept. The indicator, which characterized the city as a socio-economic system more meaningfully, was prioritized.

The inflation was counted by recalculation in the prices of 2015 based on consumer price indexes, different population of cities was taken into account on the transition to relative indicators per capita. The small sample size makes it possible to use the system as a Gaussian random vector.

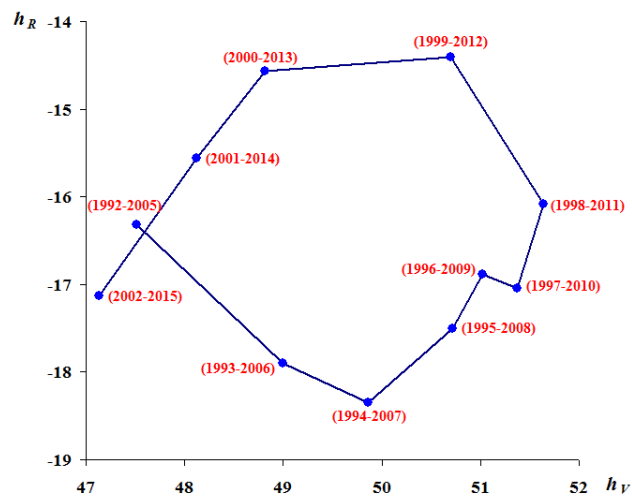
For the entropy calculations the assessments were done for periods of 13 years. This period proved to be optimal by mean of statistical smoothing, on one hand, and considering the dynamics of entropy change, on the other. The entropy was assessed in vector form (4).

Since the sample was quite small, the deviations of the empirical distributions of the features considered from the normal distribution cannot be practically established. Therefore, when calculating the randomness and self-organization entropies, we'll use the formulas (3).

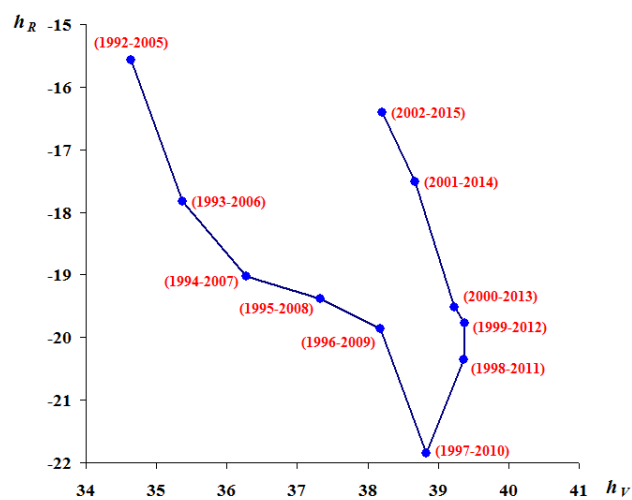
In Figure 1–3 there are the plots of changes in the randomness and self-organization entropies in Ekaterinburg, Moscow and St. Petersburg, respectively.



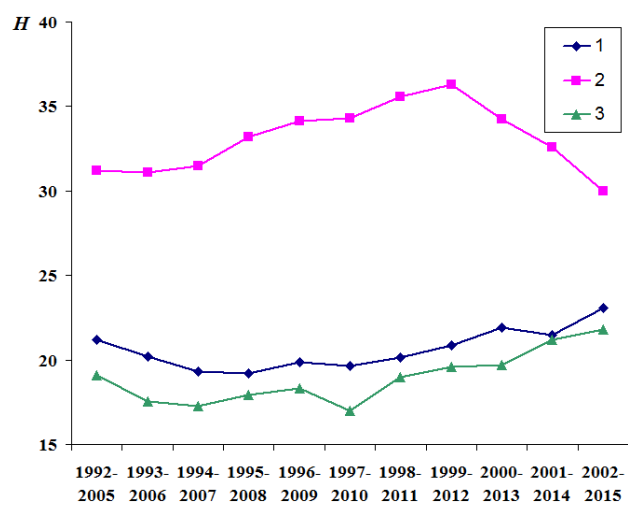
**Figure 1.** The randomness and self-organization entropy changing in Ekaterinburg



**Figure 2.** The randomness and self-organization entropy changing in Moscow



**Figure 3.** The randomness and self-organization entropy changing in St. Petersburg



**Figure 4.** Dynamics of entropy in: 1 – Ekaterinburg, 2 – Moscow, 3 – St. Petersburg

Figure 4 shows the dynamics of entropy (2) for all three megacities.

The analysis of Figure 1–4 plots according to the entropy model allows us to draw the following conclusions.

- 1) Initially, Moscow was at a higher level of development, because it had a greater variety of infrastructures functioning (the greatest value of randomness entropy) and, at the same time, demonstrated a higher level of interaction between infrastructures (minimum value of self-organization entropy). St. Petersburg was the second, both in terms of the diversity of the infrastructures functioning, and of the interaction between them.
- 2) Ekaterinburg was the third place in the initial period. But during the analyzed period, Ekaterinburg showed the best dynamics of development and almost equaled with St. Petersburg in levels of randomness and self-organization entropies. Moscow as a multidimensional stochastic system showed no development. The plot of Figure 2 shows a cycle, i.e. the system returned to its original state.
- 3) The graphs clearly show the specific areas of instability, corresponding to the global financial crisis of 2008–2009, the period of slower growth in gross domestic product in 2012 and the period of sanctions, since 2014.
- 4) From the graphs in Figure 4 we see that the total entropy in all three megacities has practically not changed during the examined period. This speaks about the need of considering the entropy of a multidimensional stochastic system, not in scalar, but in vector form.

### *3.2. Comparative entropy analysis for the environmental protection situation of the regions of Ural Federal District*

Entropy analysis can be carried out both at the level of a megacity or region, and at the level of their subsystems (elements), for example, critical infrastructures.

The problem of environmental protection is now becoming more acute. Therefore, as an example of the possibilities of entropy analysis for infrastructures, we can consider the environment. In the statistical reporting of the Federal State Statistics Service, the most complete data are presented for regions. We will perform a comparative entropy analysis of the state of environmental protection for four regions of the Ural Federal District – Kurgan, Sverdlovsk, Tyumen and Chelyabinsk.

The system of attributes will be formed and entropy calculation will be carried out the same way as in example 3.1.

The analysis will be carried out on the statistical data of Federal State Statistics Service for 2005–2015 [14].

From the variety of the indicators that characterize the environmental protection a system of 9 features is formed:

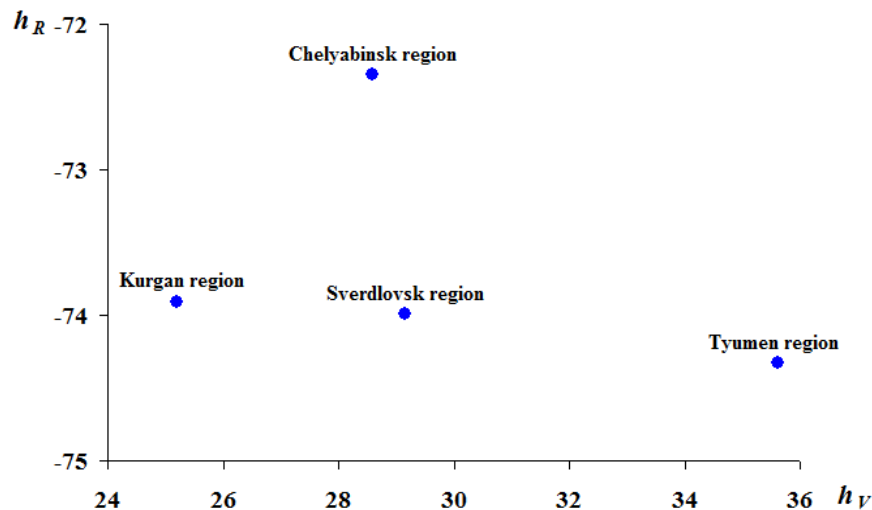
- 1) Investments in fixed assets aimed to the environmental protection, rubles per capita;
- 2) Current costs of environmental protection, rubles per capita;
- 3) Emitted pollutants of the atmosphere, kg per capita;
- 4) Water abstraction from natural reservoirs, m<sup>3</sup> per capita;
- 5) Loss of water during transportation, cc
- 6) Sewerage discharge/ contamination of surface water, m<sup>3</sup> per capita;
- 7) Area of urban settlements, that are cleaned by mechanized method, m<sup>2</sup> per capita;
- 8) The total area of the forest funds and other categories of land where forests are located, ha per capita;
- 9) The area of dead forest stands, ha per capita.

Just as in the example 1, the inflation was counted by recalculation in the prices of 2015 based on consumer price indexes; the different population of regions was taken into account by the transition to relative indicators per capita. The small sample size makes it possible to use the system as a Gaussian random vector.



In Figure 5 we have the randomness and self-organization entropies of the environmental protection system of the Kurgan, Sverdlovsk, Tyumen and Chelyabinsk regions in 2005–2015. The results of entropy analysis speak about the following.

The most developed is the environmental protection system of the Tyumen region (the largest value of the randomness entropy and the lowest value of the self-organization entropy), the next one is the Sverdlovsk region, and the Kurgan and Chelyabinsk regions are the third and fourth places, respectively.



**Figure 5.** The randomness and self-organization entropies of the environmental protection system of Kurgan, Sverdlovsk, Tyumen and Chelyabinsk regions in 2005–2015 years

#### 4. Conclusion

1. Hereby it is considered the possibilities of using the entropy modeling for the problems of sustainable development of large cities and regions, as well as the critical infrastructures on the example of environmental protection.

2. It is shown that the vector representation of the system entropy allows us to reveal the main trends in the development of multidimensional stochastic systems.

#### Acknowledgements

The reported study was carried out by financial support of the Russian Fund of Fundamental Research grant No. 17-01-00315 “Development of models and methods of monitoring, forecasting and control optimization of multidimensional regional socio-economic systems based on entropy and minimax approaches”.

#### References

- [1] O'Sullivan A 2012 *Urban Economics* 8<sup>th</sup> edition (McGraw-Hill/Irwin)
- [2] Satzinger J W, Jackson R B and Burd S D 2008 *Systems Analysis and Design in a Changing World* 5th edition (Course Technology)
- [3] Popkov Yu S 2014 *Mathematical Demoeconomics: Integrating Demographic and Economic Approach* (Berlin: De Gruyter)
- [4] Nicolis G and Prigogine I 1977 *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (New York: Wiley)
- [5] Lindblad G 1983 *Non-Equilibrium Entropy and Irreversibility* (Dordrecht: Delta Reidel)
- [6] Haken H 2000 *Information and Self-Organization: A Macroscopic Approach to Complex Systems* (Springer)
- [7] Ellis R S 2006 *Entropy, Large Deviations, and Statistical Mechanics* (Springer)



- [8] Timashev S A and Tyrsin A N 2011 Entropy Approach to Risk-Analysis of Critical Infrastructures Systems *Proc. First Int. Conf. on Vulnerability and Risk Analysis and Management and Fourth Int. Symposium on Uncertainty Modeling and Analysis* (Hyattsville, Maryland: ASCE) p 147–54
- [9] Wilson A G 1970 *Entropy in Urban and Regional Modelling* (London: Pion)
- [10] Webber M J 1979 *Information Theory and Urban Special Structures* (London: Croom Helm)
- [11] Shannon C E 1948 A Mathematical Theory of Communication *The Bell System Technical Journal* **27** pp 379–423, 623–56
- [12] Tyrsin A N 2016 *Entropy modeling of multidimensional stochastic systems* (Voronezh: Scientific Book)
- [13] *Regions of Russia The main social-economic indicators of cities* Federal State Statistics Service Available from: [http://www.gks.ru/wps/wcm/connect/rosstat\\_main/rosstat/en/statistics/publications/catalog/](http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/statistics/publications/catalog/)
- [14] *Environmental protection in Russia* Federal State Statistics Service Available from: [http://www.gks.ru/wps/wcm/connect/rosstat\\_main/rosstat/en/statistics/publications/catalog/](http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/statistics/publications/catalog/)