STATISTICAL MODELLING OF BEHAVIOUR OF URBAN ECO-ENVIRONMENT SYSTEM

Vladimirs Jansons, Vitalijs Jurenoks, Konstantins Didenko
Riga Technical University, Latvia

Abstract

Statistical modelling is one of the most widespread methods of research of economic and urban eco-environment systems. The selection of methods of modelling of the urban eco-environment systems depends on a great number of conditions (modelling components) of the system being researched. The method of statistical modelling allows developing different scenarios of functioning of the investigated economic and urban eco-environment systems. Most of the economic processes and urban eco-environment systems are complex entities, consisting of a great number of interrelated subsystems (which in their turn also are complex objects and require a detailed study), changing their positions in space and time. For researching economic and urban eco-environment systems it is impossible to create an effective model by applying traditional analytical methods. In such cases it is necessary to use the methods of statistical modelling using Monte Carlo method. In the process of modelling the most frequently method used to model multivariate distribution incidental values is the parametric method of modelling. In this case it is necessary to establish parameters of common distribution of incidental values characterizing the factors under consideration. Usually this is done by means of evaluation of parameters of multivariate distribution, i.e. by establishing the most suitable distribution, deriving from the available empirical data. On the basis of the obtained model it is possible to estimate the behaviour of the investigated economic and urban eco-environment systems in relation with probabilities and therefore its expected values what is not possible to do with classical methods. When establishing the distribution of parameters describing the behaviour of investigated economic and urban eco-environment systems from empirical information most frequently is insufficient for a credible assessment of parameters offered by the function of distribution. In these cases it is necessary to use nonparametric modelling methods, given distribution of incidental values and then modelling parameters of distribution. The objective of the paper is to describe the technique of using of statistical modelling methods for investigation of economic and urban eco-environment systems.

Key words: Statistical modelling, economic and urban eco-environment systems, effectiveness

Statistical modelling is one of the most widespread methods of research of economic and urban eco-environment systems. The selection of methods of modelling of the urban eco-environment systems depends on a great number of conditions (modelling components) of the system being researched. The method of statistical modelling allows developing different scenarios of functioning of the investigated economic and urban eco-environment systems. Most of the economic processes and urban eco-environment systems are complex entities, consisting of a great number of interrelated subsystems (which in their turn also are complex objects and require a detailed study), changing their positions in space and time. For researching economic and urban eco-environment systems it is impossible to create an effective model by applying traditional analytical methods. In such cases it is necessary to use the methods of statistical modelling using Monte Carlo method. In the process of modelling the most frequently method used to model multivariate distribution incidental values is the parametric method of modelling. In this case it is necessary to establish parameters of common distribution of incidental values characterizing the factors under consideration. Usually this is done by means of evaluation of parameters of multivariate distribution, i.e. by establishing the most suitable distribution deriving from the available empirical data. On the basis of the obtained model it is possible to estimate the behaviour of the investigated economic and urban eco-environment systems in relation with probabilities and therefore its expected values what is not possible to do with classical methods. When establishing the distribution of parameters describing the behaviour of investigated economic and urban eco-environment systems from empirical information most frequently is insufficient for a credible assessment of parameters offered by the function of distribution. In these cases it is necessary to use nonparametric modelling methods, given distribution of incidental values and then modelling parameters of distribution. The main objective of the paper is to describe the main ideas of using of statistical modelling methods for investigation of economic and urban eco-environment systems.

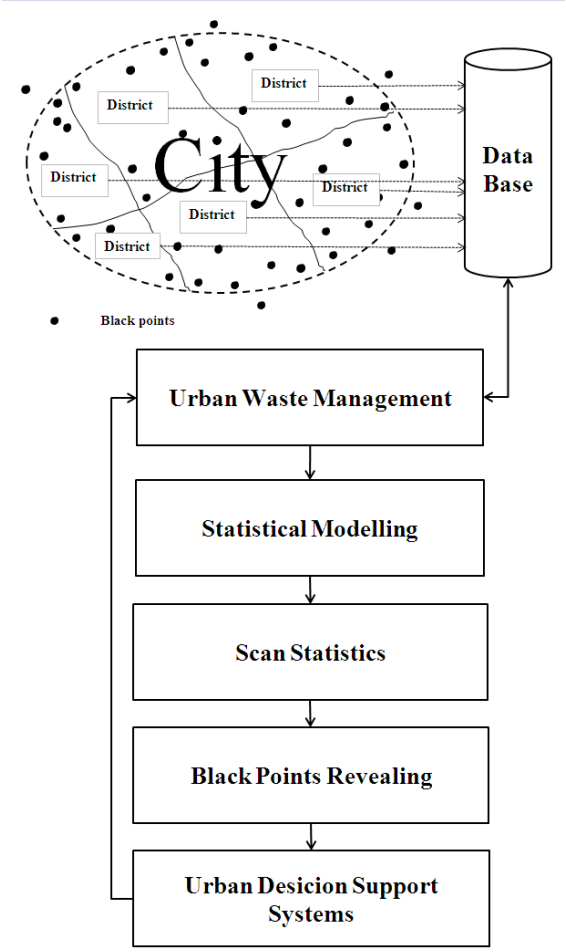

Fig. 1. Urban waste management structure – urban statistical information system

Main role in waste management takes information technologies. The modern degree of development of information systems in social and economic area allows solving the problem on revealing factors which influence development urban social-economic systems as a whole. These allow an opportunity to use methods of statistics (scan statistics methods) more widely, using greater files of the information on dynamics of development of investigated object for different social, economical points of view. For example, a growing urban activity implies appearances of the extra-waste (black points) in no predictable place of the city street structure. Scan statistics method allows quickly and accurately determine whether is occurring. Thus the main problem is quickly and accurately determines whether the extra-waste is occurring. The use of centralized data recording and databases generates large quantities of data, with consequent limitations on human resources to search and analyze these data for waste (black points) clustering. Automated processes to evaluate waste clusters in place and in time (temporal clustering) are needed to enhance the efficiency of urban waste management. Detection of temporal clustering by using of the scan statistics offers some advance towards rapid detection of extra-waste, and modern statistical modeling methods have been made available.

In most cases possibility of traditional statistical conceptions and methods for investigation of real socio-economical object is bounded. Traditional statistical methods are more appropriate for investigation of influences of local factors – for localized objects investigation. Scan statistics allows investigating the socio-economical problems having extremely complex urban socio economical structure. The analytical description of such systems probably only in the simplified kind, enabling to consider the likelihood scenarios of development of investigated object, but it is rare when all object as a whole.

Methodology of using of scan statistics for waste clustering.

In each area Z, we assume that the data X about waste level have a distribution function DF (to be distributed under null hypothesis H_0) i.e.:

$$X \sim DF \tag{1}$$

We also compute the maximum of L_1, which is the same function with parameters unrestricted. Each zone Z has different parameters, given the heterogeneous accident distribution. We want to find the zone which maximizes the LR (likelihood ratio) between likelihoods L_1 and L_0:

$$LR(Z)=\left(\frac{L_1}{L_0}\right)_Z \tag{2}$$

In the case of Poisson distribution process, the likelihood ratio takes the following form:

$$LF_i=\left(\frac{\left(\frac{c_{in}}{n_{in}}\right)^{c_{in}}\left(\frac{c_{out}}{n_{out}}\right)^{c_{out}}}{\left(\frac{c_{tot}}{n_{tot}}\right)^{c_{tot}}}\right)\cdot I\,, \tag{3}$$

where:

c - waste level;

n - expected level of waste;

n_{in} and n_{out} - within or outside the scanning window;
I - indicator function.

The scan statistics LR_{st} is defined as:

$$LR_{st} = \max_{Z} LR(Z) \quad (4)$$

Null hypothesis significance testing.

For each potential cluster, we generate N datasets using the parameters θ_0 estimated for that zone Z, and we obtain a distribution for LR. Later the distribution LR is used for identifying the statistical value of the cluster detected. The scheme for identifying a significant cluster using the Monte-Carlo method is presented in Figure 2.

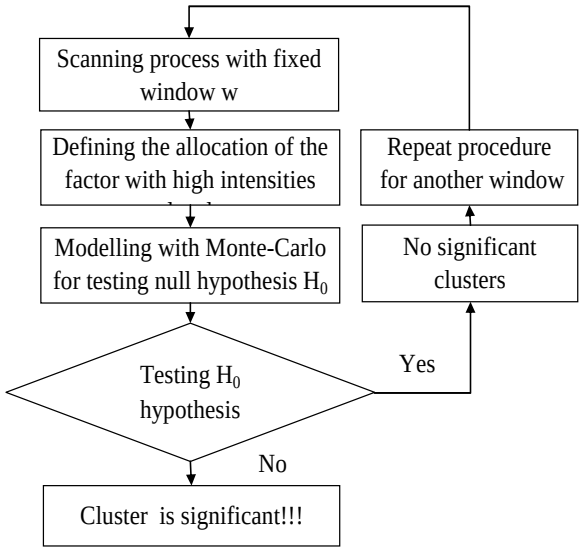

Fig. 2. Algorithm of defining a significant waste cluster

The algorithm developed enables to detect the statistically significant clusters of the phenomena under investigation.

Results of generated N datasets of distribution for LR using the parameters θ_0 are shown in Figure 3:

Spatial scan statistics is a powerful method for spatial cluster detection. With spatial scan statistics it is possible to search over a given set of spatial regions, find those regions which are most likely to be clusters and correctly adjust for multiple hypothesis testing.

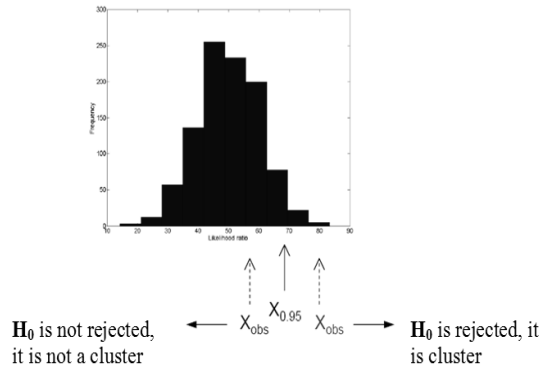

Fig. 3. Histogram for ratio test statistic

Figure 4 illustrates a suspicion cluster – region in S with a high level of intensity q_{in} = 0.02 of accidents. Scan statistics gives answer to the question – is this waste cluster real or is it a "visual illusion"?

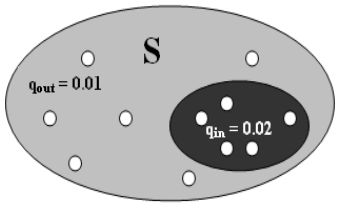

Fig. 4. Frequency model of cluster - critical region

The simplest frequency model for this situation (Figure 4) can be written as:

- null hypothesis H_0 (no clusters) $q_{in} = q_{all}$ everywhere (use maximum likelihood estimate of q_{all});
- alternative hypothesis H_1 (cluster in region S), $q_{in} = q_{out}$ elsewhere (use maximum likelihood estimates of q_{in} and q_{out}, subject to $q_{in} > q_{out}$).

This algorithm can be used for scanning accidents on the territorial unit of Latvia (Figure 5).

The spatial scan statistics have been used to detect and extract spatiotemporal clusters of wastes within the districts of city. As a case study, the distribution of wastes and logistic problems has been considered so far. Results show a dependency between high- and low-rates districts: in the low-rates districts black points number significantly less than in the high-rate districts. The use of district degree of technology and population density allows understanding the reason of the waste (black points) clustering in a certain district area. Results of this transdisciplinary research will increase the understanding of urban phenomena; improve urban waste man-

agement analysis and modelling waste management processes.

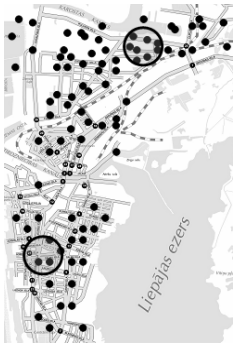

Fig. 5. Illustration of the scanning process of accidents on roads of Latvia (small territorial fragment)

In the Figure 6 black point's density development in some region is illustrated. After 3 years we see the relocation of the maximum black point's intensity factor from one sector to the other sector of the district. The modern informational technology allows improving data collection and data management processes. Scan statistics appears naturally at attempt to define clusters of events in our case waste clusters (black points), using the saved up statistical information about ecological situation (waste distribution by places and by time, accidental points – spot points, noise level in every district).

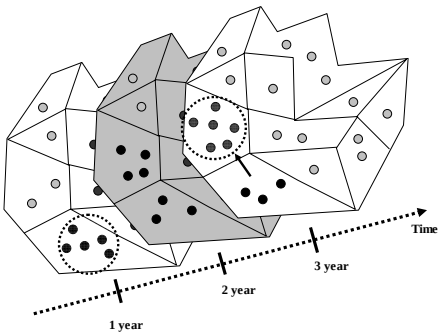

Fig. 6. Spatial-time scan statistic using for some city district.

Waste management problem fast increasing last time. Modelling different mixed strategies for different waste groups we can find more appropriate (cheapest way) strategy for all waste recycling (see Figure 7).

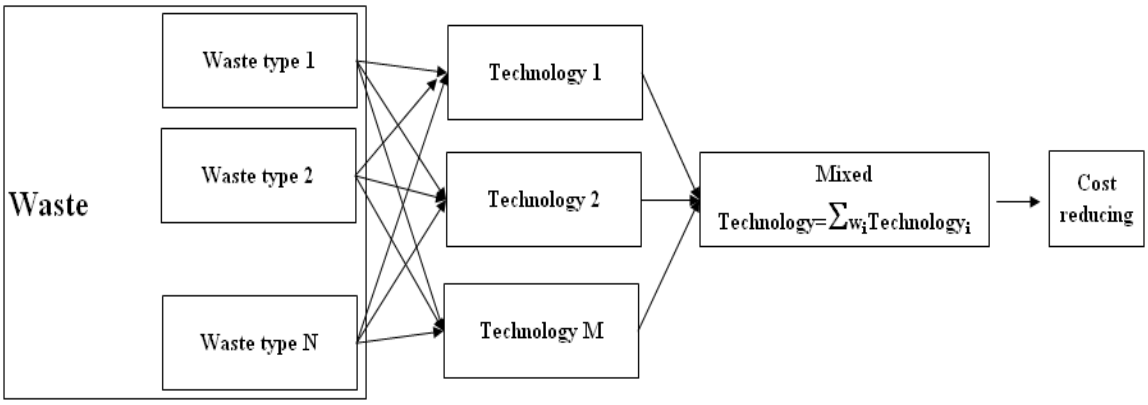

Fig. 7. Waste recycling mixed strategies for waste management cost reduction.

The aim of modelling waste recycling mixed strategies is to ensure the recovery and recycling of packaging waste in the most economically efficient and ecologically sound manner.

Conclusion

The scan statistic methodology at this stage looks at scenarios and probability models for the null distribution in the time series of events, assuming a constant background rate of random events according to the null distribution. It is important to distinguish the appropriate model for the relevant sampling process. Two major concerns with existing methods or approximations of a scan statistic are, knowing what really is correct, and what assumptions are made about statistical distributions used in the approximation. Scan statistic has wide area of applications: marketing, community infrastructure, disaster management, ecosystem health, air pollution, waste management, robotic networks, environmental management and policy, public health and environment, social networks.

The use of scan statistics enabled:

- to make analysis of waste management in towns and regions of Latvia;
- to detect clusters with utmost waste intensity applying scan windows of different sizes;
- to check significance of clusters detected with highest frequency of waste (black points) on the basis of null hypothesis equal to value of 0,05;
- to analyse the dynamics of changes of clusters detected taking into consideration the time factor.

Detection of significant clusters of waste (black points) in towns and regions of Latvia enables to take prompt actions for improving the quality of waste management in Latvia.

Reference

1. Jurenoks, V., Jansons, V., Didenko, K., *Modelling of financial stability in logistics in conditions of uncertainty*, 21-st European Conference on Modelling and Simulation, Prague, Czech Republic, June 4 – 6, 2007
2. Jurenoks, V., Jansons, V., Didenko, K., *Modelling of multidimensional flows in logistics using non-parametric method*, International Mediterranean Modelling Multiconference, Bergeggi, Italy, October 4 – 6, 2007
3. Kulldorff, M. *A spatial scan statistics. Communications in Statistics. Theory and Methods*, 1997, Volume 26, Issue 6, 1481 – 1496
4. Glaz, J.; Naus, J., Wallenstein, S. *Scan Statistics. Springer Series in Statistics*, 2001, 367-370
5. Wallenstein, S. *Scan Statistic*, 2005, http://c3.biomath.mssm.edu/wscan.html.
6. Ward, M.P. and Carpenter, T.E. *Methods for Determining Temporal Clusters in Surveillance and Survey Programs. In: Animal disease Surveillance and Survey Systems. Methods and Applications.* Ed. Salman, M.D. Iowa State Press, Ames Iowa, 2003, p 87-99
7. Tuia, D., Kaiser, C., Da Cunha, A., Kanevski, M. *Socio-economic cluster detection with spatial scan statistics. Case study: services at intra-urban scale*, Geocomputation 2007, National University of Ireland, 2007, Maynooth, 3-5 September 2007, http://www.clusterville.org/?CaseStudies/ScanStat
8. Martin Kulldorff. *A spatial scan statistic. Communications in Statistics - Theory and Methods*, Volume 26, Issue 6 1997 , pages 1481 – 1496
9. Tuia, D., Kaiser, C., Da Cunha, A., Kanevski, M. *Socio-economic cluster detection with spatial scan statistics. Case study: services at intra-urban scale*, Geocomputation, National University of Ireland, Maynooth, 3-5 September 2007
10. Ward, M.P. and Carpenter, T.E. *Methods for Determining Temporal Clusters in Surveillance and Survey Programs. Animal disease Surveillance and Survey Systems. Methods and Applications. Ed. Salman, M.D*, Iowa State Press, Ames Iowa, 2003, 87-99