INVESTIGATION OF ACCIDENT AND NOISE INTENSITY ON LATVIAN ROADS USING SCAN STATISTICS

Vladimirs Jansons[a], Vitalijs Jurenoks[b], Konstantins Didenko[c]

[a, b, c]Riga Technical University, Faculty of Engineering Economics, Kalku iela 1, Riga, LV1658, Latvia

[a]vladjans@latnet.lv, [b]vitalijs.jurenoks@rtu.lv, [c]konstantins.didenko@rtu.lv

ABSTRACT

In most cases the possibility of traditional statistical conceptions and methods for investigation of real socio-economic objects is limited. Traditional statistical methods are more appropriate for investigation of impact due to internal factors. Scan statistics allows investigating the socio-economic problems having extremely complex, i.e. synergic structure of interrelations (structure of the open systems). The analytical description of such systems enables to consider the likelihood scenarios of development of objects under investigation in a simplified way, but very seldom of objects on the whole. The use of scan statistics enables to understand and analyse the reasons of the clusters of accident and noise intensity in some specific areas of Latvia.

Keywords: modelling, scan statistics, clustering, black spots, intensity of noise, road system management

1. INTRODUCTION

With the increasing road transport vehicle saturation on highways the role of road transportation network monitoring becomes more and more significant. Taking into consideration that Latvia is a transit country with a heavy road traffic, reduction of noise level on the highways of Latvia and control of the physical condition of roads has become the most urgent problems. For several years measurements of noise level intensity on roads have been made and maps of highway noise spreading near densely populated areas are compiled. Although full information of road accidents and high noise intensity level on highways is not yet amassed, in some cases it has already become possible to highlight some clusters of road accidents and noise intensity. Gathering of information and its intermediate processing will contribute to identification of clusters, so that prompt decisions could be taken. Statistical modelling is used to state the clusters of accidents and level of noise intensity on the roads of Latvia. Alongside with the Monte-Carlo method when calculating the p-value, it is possible to use the direct method of scan modelling, taking into consideration the empirical rules of distribution of road accidents on the highways of Latvia. In this case both parametric and non-parametric methods may be used to describe the distribution of road accidents and noise intensity in Latvia. The research data may be used for developing an efficient strategy of road system management in Latvia leading to reduction of the number of road accidents and noise intensity in Latvia.

2. METHODOLOGY OF SCAN STATISTICS

In each area Z, we assume that the data X about accidents or noises 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 - counts;
n - expected number of cases;
n_{in} and n_{out} - within or outside the scanning window;
I - indicator function.

The scan statistics LR_{st} is defined as:

$$LR_{st}=\max_Z LR(Z) \,. \tag{4}$$

2.1. Null Hypothesis Significance Testing

For each potential cluster, we generate N datasets (N> 1000) using the parameters λ_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 1.

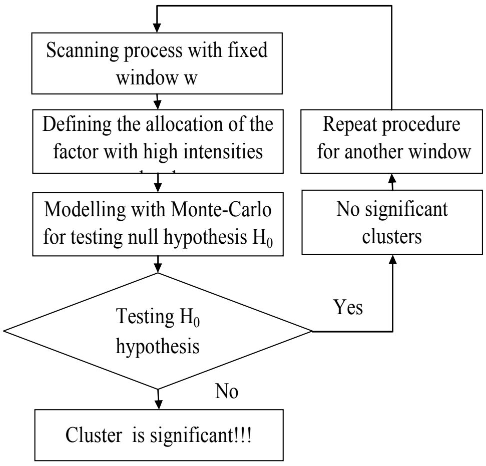

Figure 1: Algorithm of defining a significant cluster

For each potential cluster, we generate N datasets (at least 1000) using the parameters λ_0 estimated for that zone, and we obtain a distribution for LR (see Figure 2):

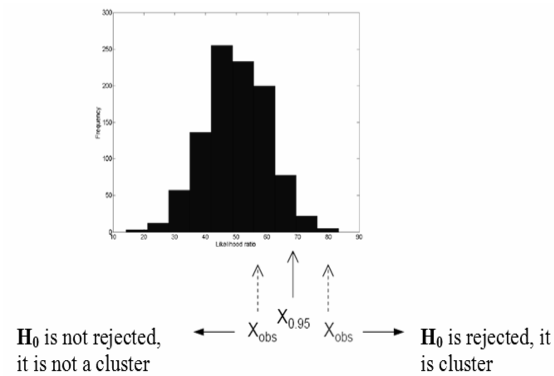

Figure 2: Histogram for ratio test statistic

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

Let us have N accidents, distributed at road interval (0, L). Denote S_w as maximal number of events at a road interval with length w (the window of fixed length w). The maximum cluster S_w is called the scan statistics from the viewpoint that one scans interval (0, L) with a window of size w and observes a large number of points. W_k is the shortest interval of road containing a fixed number of k events. The interval W_{r+1} is called the minimum r^{th} order gap, or r-scan statistics. The distributions of the statistics S_w and W_k are related. If the shortest window that contains k points is longer than w, then there is no window of length w that contains k or more points:

$$P(W_k > w) = P(S_w < k). \tag{5}$$

It is easy to understand that there is an infinite number of sliding windows during the road interval. To solve this problem in a constructive way we must assume some finite set of sliding windows (Figure 3).

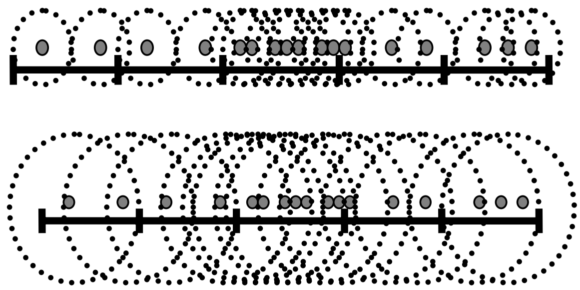

Figure 3: Illustration of the scanning window of two fixed lengths w=0.1 and w=0.2. The centres of scanning windows are points with road coordinates $L_1, L_2, \ldots, L_N$

The Poisson process has been used for modelling real systems dealing with the occurrence of events in time or space. First useful applications of spatial scan statistics are shown in "A spatial scan statistics. Communications in Statistics. *Theory and Methods*" (Kulldorff, M. 1997).

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. 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 cluster real or is it a "visual illusion"?

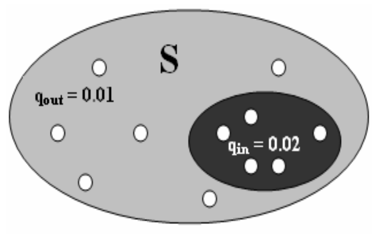

Figure 4: Frequency model of cluster - critical region

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).

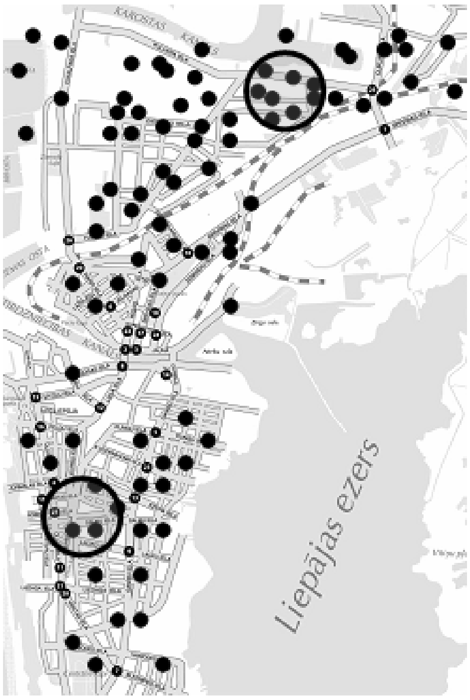

Figure 5: Illustration of the scanning process of accidents on roads of Latvia (small territorial fragment)

Preliminary results of investigating accidents on the roads of Latvia show the following specific features of their distribution:
- lack of traffic saturation in rural areas does not allow to correctly identify clusters of road accidents in these regions;
- significant clusters occur in big cities of Latvia showing possibilities of improving road management system in Latvia (see Figure 6).

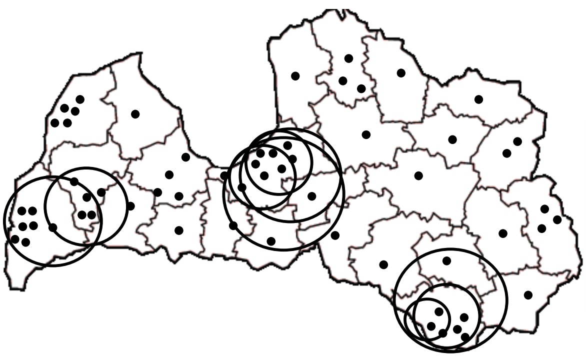

Figure 6: Two dimensional scan statistics with different circle windows

3. NOISE MAPS

Recent years have seen an increasing need to determine the noise from road traffic. This is often necessary, for example, when are implementing traffic and environmental action plans, when planning new housing estates and roads, and when residents complain about noise. Data on noise in urban environment may be obtained through a direct measurement or by means of model calculations. Both the methods (modelling and noise measurement) have their advantages and drawbacks. The measurement provides more accurate data and reflects the real noise level from all sources of noise at the locality monitored. Performance of the measurement over a larger area is, however, very financially demanding because it requires a great deal of resources to take qualified noise measurements. The uncertainty involved in measurement is roughly the same as with modelling. If noise modelling is performed, on the other hand, the noise in any given situation can be determined relatively quickly. Furthermore, the modelling method is particularly applicable in planning situations, where the noise cannot yet be measured. This is why modelling is generally used to determine the noise. The noise modelling is limited not just due to the modelling method employed, yet most of all by availability of input data. On the other hand, it enables making of qualified assessments of expected impacts of planned development and transport measures.

The most important source of excessive noise affecting the largest portion of the city population is road traffic, because number of automobiles and transported volumes in Latvia has been rapidly increasing.

The example of benchmarking process for noise level investigation in Latvia is presented in the paper (as example of "best practice" has been used the investigation and measurement of noise level in the Czech Republic).

Information on noise levels on the Prague's territory has been generated at various occasions. Number of occasional measurements and assessments are implemented in the relation to the investments and development of environmental studies. Since 1984 in Prague have been regular measurements performed at selected localities within the IOŽIP system (Prague Environmental Information System). Noise level measurement is also performed by the Public Health Authority of the City of Prague within the National programme "Monitoring of the Public Health Related Environmental Aspects".

Illustration of results of measurement of noise on roads of some Prague's district is presented in Figure 7.

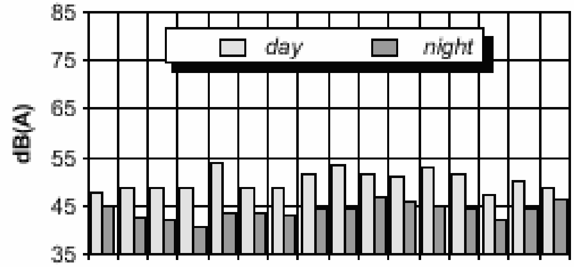

Figure 7: Illustration of results of measurement of noise on roads of some Prague's district

The objective of benchmarking is to understand and evaluate the current position of a business or organisation in relation to "best practice" and to identify areas and means of performance improvement. Any business process can be benchmarked. Benchmarking is a tool to help you improve your business process and begin comparing the results in an "apples-to-apples" format, to determine the gap between your object (process) and the best-in-class examples and to implement programs and actions for achieving needed result. Application of benchmarking in investigation process involves four key steps:

1. understand in detail existing business processes;
2. analyse the business processes of others;
3. compare own business performance with that of others analysed;
4. implement the steps necessary to close the performance gap.

Benchmarking module scheme is presented in Figure 8.

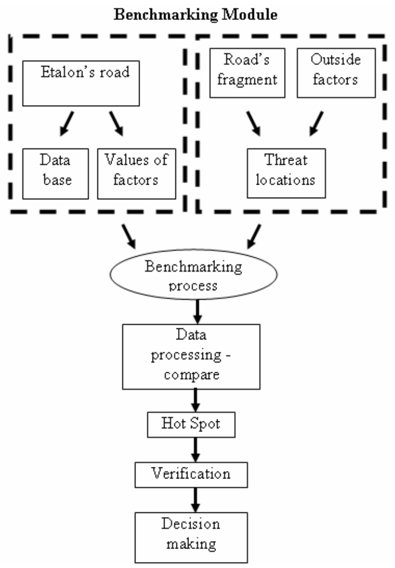

Figure 8: Benchmarking Module

Benchmarking should not be considered a one-off exercise. To be effective, it must become an ongoing, integral part of an ongoing improvement process with the goal of keeping abreast of ever-improving best practice. Detailed algorithm of realization of benchmarking process presented in Figure 9.

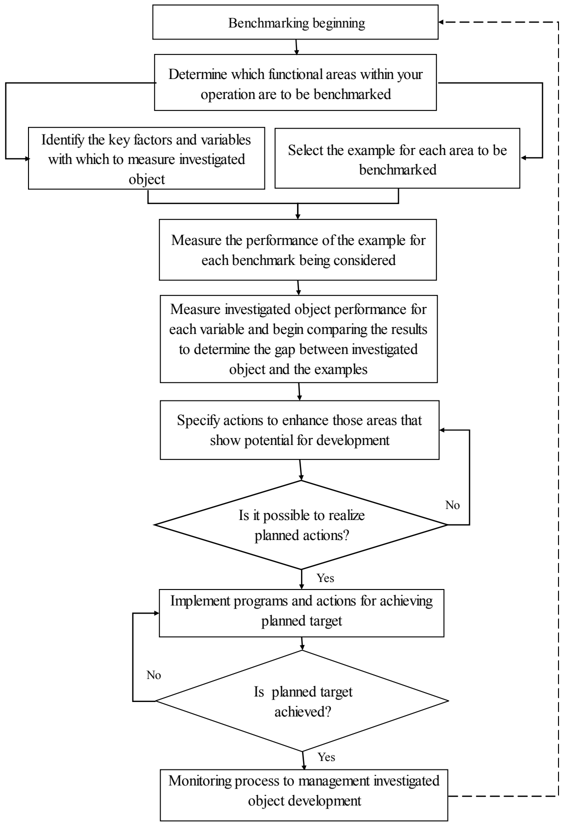

Figure 9: Algorithm of realization of benchmarking process

Accordingly with Directive 2002/49/ec of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise the level of noise is divided by 7 sublevels (intervals):

Level 1. – 15 dB;
Level 2. – 35 dB;
Level 3. – 60 dB;
Level 4. – 75 dB;
Level 5. – 85 dB;
Level 6. – 115 dB;
Level7. – 125 dB.

Illustration of results of measurement of noise on roads of Latvia is shown in Figure 10.

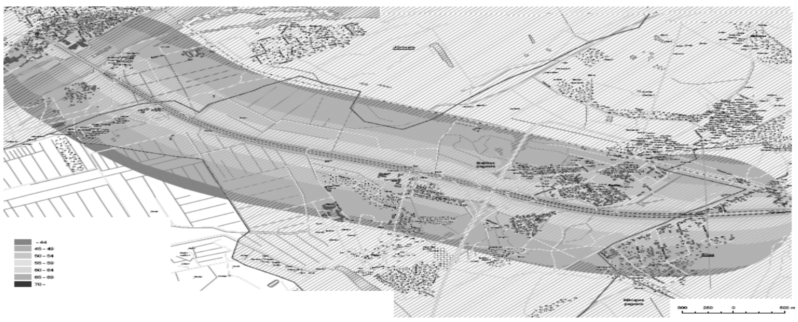

Figure 10: Illustration of results of measurement of noise on roads of Latvia (small territorial fragment)

4. CONCLUSION

The use of scan statistics enabled:

- to make analysis of road accidents and intensity of noise in towns and regions of Latvia;
- to detect clusters with utmost intensity of road accidents applying scan windows of different sizes;
- to detect clusters with utmost noise intensity applying scan windows of different sizes;
- to check significance of clusters detected with highest frequency of road accidents 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 road accidents or level of noise intensity on the roads of Latvia enables to take prompt actions for improving the quality of road infrastructure.

REFERENCES

Berman, M. and. Eagleson, G. K. Dec., 1985. A Useful Upper Bound for the Tail Probabilities of the Scan Statistics. *Journal of the American Statistical Association*, Vol. 80, No. 392, doi: 10.2307/2288548, 886-889.

Directive 2002/49/ec of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise. http://eur-lex.europa.eu/LexUriServ/

Glaz, J.; Naus, J. and Wallenstein, S. 2001. Scan Statistics. *Springer Series in Statistics*, 367-370.

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.

Jurenoks, V., Jansons, V., Didenko, K., Modelling of multidimensional flows in logistics using nonparametric method. *International Mediterranean Modelling Multiconference*, Bergeggi, Italy, October 4 – 6, 2007.

Kulldorff, M. 1997. A spatial scan statistics. Communications in Statistics. *Theory and Methods*, Volume 26, Issue 6, 1481 – 1496.

Road traffic accidents. http://www.csb. gov.lv/ csp/

Tuia, D., Kaiser, C., Da Cunha, A., Kanevski, M. 2007. 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.

Wallenstein S. and Naus J. 2003. Statistics for temporal surveillance of bioterrorism. *Syndrome Surveillance: Reports from a National Conference, Morbidity and Mortality Weekly Report.* 2004; 53, 74-78.

Wallenstein, S. 2005. Scan Statistics. http://c3.biomath.mssm.edu/wscan.html

Ward, M.P. and Carpenter, T.E. 2003. 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, 87-99.

Yearbook Prague Environment 2006 - B5.2 Road Traffic Noise. http://envis.praha-mesto.cz/rocenky/pr06_htm/ab5_02.htm.

AUTHORS BIOGRAPHY

VLADIMIRS JANSONS was born in Daugavpils, Latvia and is a graduate of the University of Latvia, where he studied mathematics and obtained his degree in 1970. For eight years he worked in the Computing Centre of the University of Latvia. Since 1978 he has been lecturing at the Riga Technical University, where in 1983 he was awarded the doctoral degree in mathematics. The main field of research pursued is simulation and optimization of economic systems.

VITALIJS JURENOKS was born in Riga, Latvia. In 1976 he graduated from the Faculty of Engineering Economics of Riga Technical University, and for ten years worked in an industrial enterprise in Riga. Since 1986 he has been lecturing at the Riga Technical University, and in 1987 was awarded the doctoral degree in economics (Dr.oec.). The main field of research pursued is planning, simulation and optimization of economic processes and systems.

KONSTANTINS DIDENKO was born in Jelgava, Latvia. In 1969 he graduated from the Faculty of Engineering Economics of Riga Technical University. Since 1969 he has been lecturing at the Riga Technical University where in 1985 he obtained the doctoral degree in economics. In 2006 he was elected a corresponding member of the Latvian Academy of Sciences. The main field of research pursued is planning and optimization of economic processes and systems