How can find the developing territories?

Karin Lindroos
Tallinn University of Technology
Kopli 101, 11712 Tallinn, Estonia
karin@tv.ttu.ee, tel +372-6204111

Ants Torim
Tallinn University of Technology
Raja 15, 12618 Tallinn, Estonia
torim@staff.ttu.ee, tel. +372-6202306

Leo Võhandu
Tallinn University of Technology
Raja 15, 12618 Tallinn, Estonia

leov@staff.ttu.ee, tel. +372-6202304

Abstract

Different regional development and inequality between the places are main problems in spatial economics. Formation of social organization and spatial hierarchy depend on some certain advantages, which are also the key factors of development.

Our goal is to find these factors and developing territories as potential central places.
We use in observable work large amount of variables, because of this we try to use a novel approach for social economic data analysis – scale of conformance and leading functions of the monotone systems. Large amount of variables makes use of standard statistical methods in spatial economic research extremely difficult.
Our methods requires discrete data table and they are based on a computationally simple weight functions that describes objects "typicality" for data table. As similar elements have similar weights, it is possible to find groups of similar settlements forming social-economical classification of the settlements.
We apply these methods for analysis of social-economical data about the functional regions, like Estonian islands Saaremaa and Hiiumaa as target areas, show that we can detect notable outliers and typical settlements, with our method. The outliers are fast developing territories, which ones we are looking for and settlements of the top position.
The scale of conformance and leading functions helps to determinate the social organization of the functional regions. Decision makers will get information about the key factors of observable local or regional spatial economics.

Keywords

Regional development, social and spatial organization, functional regions, central places, social-economic factors, monotone systems, scale of conformity, data visualization, conformity, leading function, settlement system

1. Introduction

Different regional development is main problem in spatial economics, and for that reason inequality between the places. Formation of social organization and spatial hierarchy depend on some certain advantages, which are the key factors of development.
Our goal is to find these factors and developing territories as potential central places.

We use in observable work large amount of variables, because of this we try to use a novel approach for social economic data analysis – scale of conformance and leading functions of

the monotone systems. Large amount of variables makes use of standard statistical methods in spatial economic research extremely difficult.
Our methods requires discrete data table and they are based on a computationally simple weight functions that describes objects “typicality” for data table. Use of two slightly different weight functions allows us to create two-dimensional conformity plots visualization for multivariate social-economical data. As similar elements have similar weights, it is possible to find groups of similar settlements forming social-economical classification of the settlements.
We apply these methods for analysis of social-economical data about the functional regions, like Estonian islands Saaremaa and Hiiumaa as target areas, show that we can detect notable outliers and typical settlements, with our method. The outliers are fast developing territories, which ones we are looking for and settlements of the top position.

Data in our table is binary, where meaning of some values are “existent” (one) and “missing” (zero). It is also computationally fast method where only one pass through data table is needed. We find a weight called conformity for each object in a data table. Conformity for an object is calculated by a transformation where instead of the attributes value we use its frequency (so-called frequency transformation). The most attributes are binary by nature such as existence of a port or a school. Each numerical attribute was replaced by several attributes that represent an interval. For example number of children in a village is represented by four binary attributes children <10, children 10-50, children 50-100, children >100 etc.

Use of two slightly different weight functions allows us to create novel two-dimensional conformity plots visualization for multivariate social-economical data. As similar elements have similar weights, it is possible to find social-economical groups or clusters of similar settlements Target areas have four essential groups according to scale of conformance.
A: The most non-typical villages, people do not live there.
B: Villages with no children.
C: Large settlements and administrative centres.
D: Main group, which can be divided into three subgroups, with most typical places in Estonian islands.

Leading function based also on the conformance scale, but non-typical settlements eliminate from the table according to their weights. Non-typical means here settlements with high values or places we are looking for.
Using the algorithms of leading functions are enable to separate all active places and make hierarchical system for the settlements. Viable ones pull together and push weaker villages to the outside. They form some another groups. In the plot visualization is clear difference between viable active settlements and others with low social characteristics.
Leading function also can clearly differentiate the outliers as potential new leaders (here developing settlements, new or potential central places), since in the pure conformance scale it was not so clear.

Same way we can find the key factors which have great influence to development of the places.

The scale of conformance and leading functions helps to determinate the social organization of the functional regions. Decision makers will get information about the key factors of observable local or regional spatial economics.

2. Body of paper

2.1 Overview of methods

2.1.1 Scale of conformity

We describe here a scale of conformity approach that is one of the simplest monotone systems methods. It is also computationally fast method where only one pass through data table is needed. We find a weight called conformity for each object in a data table. Conformity for an object is calculated by a transformation where instead of the attributes value we use its frequency (so-called frequency transformation). For every row in the data table we calculate the sum of all attribute-value frequencies. This sum is the conformity weight for that row. Intuitively conformity describes objects "typicality" for entire data table (system). If we include frequencies of missing and negative values (zeros in binary data table) in our conformity calculation then we are using weight function π_{01}. If we don't include frequencies of zero values (we are using only frequencies of ones in binary data table) calculation then we are using weight function π_1.

For example let us consider following data table:

Table 1. j * i binary data table

j / i	**1**	**2**	**3**	**4**	**5**	**6**
1	0	0	0	0	1	0
2	0	0	1	1	1	0
3	1	0	0	0	1	0
4	0	0	0	1	0	1
5	0	1	0	0	1	0
6	0	1	0	0	1	0
7	0	0	0	0	0	1
8	0	1	0	0	1	0

After calculating frequencies and weights we get:

Table 2. Weights and frequencies for previous table, rows are sorted after π_{01}

j / i	**1**	**2**	**3**	**4**	**5**	**6**	$\pi_{01}(j)$	$\pi_1(j)$
1	0	0	0	0	1	0	39	6
5	0	1	0	0	1	0	35	9
6	0	1	0	0	1	0	35	9
8	0	1	0	0	1	0	35	9
3	1	0	0	0	1	0	31	7
7	0	0	0	0	0	1	29	2
2	0	0	1	1	1	0	27	9
4	0	0	0	1	0	1	25	4
f (i, 0)	7	5	7	6	2	6		
f (i, 1)	1	3	1	2	6	2		

Such ordering of data table makes it possible to detect frequent itemsets visually from data table. For example itemset {i3=0, i4=0, i5=1, i6=0} with support 5 is clearly visible from our sorted table.

Data table about Hiiumaa is sparse – only 4.7% of values are ones. When using weight function π_{01} on sparse data table mostly empty rows tend to have highest weights. In this article we propose using both weight functions - π_{01} and π_1 - for data mining and visualization.

This method can be used for any discrete data table regardeless of number of attributes (dimensions). Our proposed conformity plot visualization is similar to clustering visualizations like Kohonen nets and nonlinear projection visualizations like Sammon plots [Hoffman and Grinstein, 2002]. Our visualization displays clusters and outliers. Furthermore: both axes in our visualization have intuitive meaning as they show objects typicality for data table. Most typical objects are located at upper right corner of the plot.

2.1.2 Leading functions

Two different leading functions based also on the conformance scale, but elements or settlements eliminate from the table according to their weights. Additionally we are using here two mathematically different functions (algorithms) (Muchnik, Kuznetsov, 1982).
The essential difference between these functions is following:
(1) Administrative activeness is measured by weight function, where the weight is assigned according to such functions that are not represented in other settlements.
(2) Incompetence function assigns the weight to the settlement by subtracting the number of functions that are performed by the settlement from the overall functions that are performed by settlements in general.

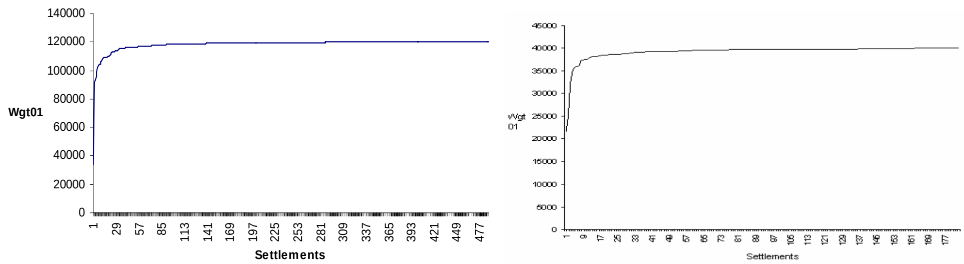
2.1 Our data

Saaremaa and Hiiumaa are the islands of the Baltic Sea with territories of 2922 and 1049 km². These islands have population in accordance with 38200 and 11000 inhabitants. Our data table contains settlements (488 in Saaremaa and 185 in Hiiumaa) and their demographic and economic characteristics (or some activities or values, accordingly 254 and 226). It is available from [Juurikas and Torim, 2006].
Data in our table is binary. Each numerical attribute was replaced by several attributes that represent an interval. For example number of children in a village is represented by four binary attributes children<10, children10-50, children50-100, children>100.

2.2 Data analysis using scale of conformity

2.2.1 Analysis of data using weight-function π_{01}

The most of the settlements have not any economic activities or functions, only social characteristics. Frequent appearance of zeros determines weights of the elements. Villages, which have few variables with value 1, have greater π_{01} weights. Observable islands have most of that kind of the settlements.



a. Saaremaa b.Hiiumaa

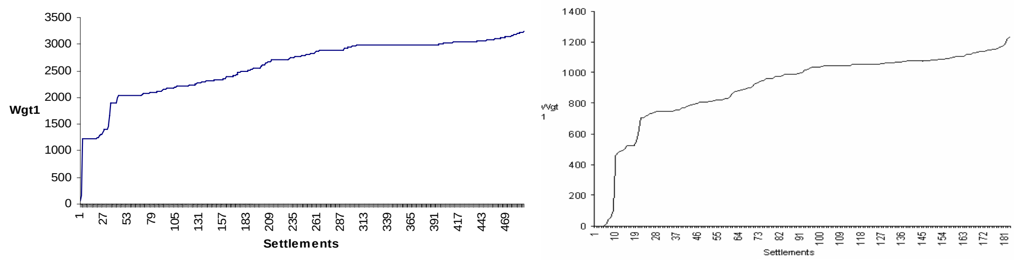
Figure 1. Settlements and their weight π_{01}, sorted in ascending order.

According to weight function π_{01}, most typical are living and summerhouse villages. The greatest weight in the scale of conformity belongs to 87 living villages in Saaremaa (for example Levala, Saue-Putla, Väljaküla, Undimäe, Lööne etc, weight 120611) and ten living villages in Hiiumaa (for example Laaritsa, Lepiku, Ulja, Aadma, Jõeküla, Ühtri, Otste, Kalgi, Pilpaküla and Sakla, weight 40029). They are the most typical elements for the system and they have also identical demographic and social variables that are most widespread in this county.

We can see from Figures 1 (a., b.), that small number of settlements have notably lower weights than others (<115000 and 36000). These are large settlements and administrative centers. They have economic activities, administrative importance and better social characteristics (more habitants, more children etc.). Their lower weights are caused by having lots of characteristics that are atypical for more common, smaller settlements. So we can see that, scale of conformity detected both typical elements and outliers.

2.2.2 Analysis of data using weight-function π_1

When weight function π_{01} is calculated on our sparse data, high frequencies of zeros tend to dominate. Weight function π_1 is calculated using only frequencies of ones. So objects are "typical" (have high weight) when they have lots of common characteristics and having uncommon characteristics does not reduce objects weight. That will give us somewhat different ordering.



a. Saaremaa b. Hiiumaa

Figure 2. Settlements and their weight π_1, sorted in ascending order.

As we can see in the Figure 2 (a., b.), settlements are divided according to the weights into the three groups:

I group, weight <1228 in Saaremaa and weight <400 in Hiiumaa. The most non-typical villages, people do not live there and villages have not any social characteristics. But they have some economic activities, like harbour, custom, border guard, etc, which are supervised from other (central) places.

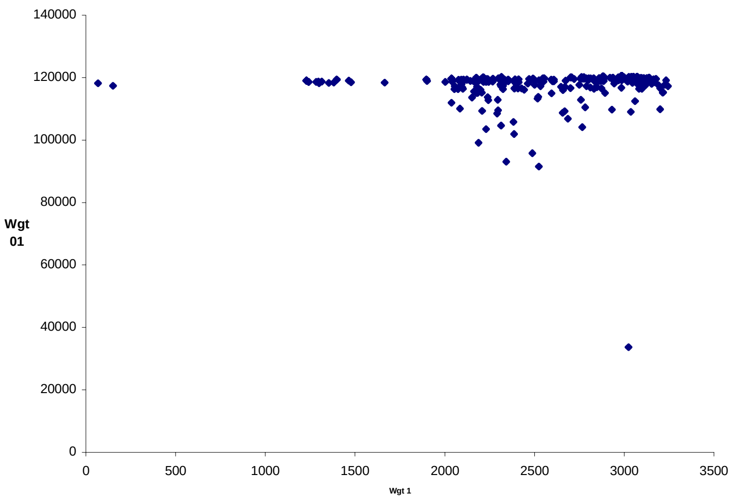
II group, 1228 < weight < 2003 (2037) in Saaremaa, 400 < weight < 700 in Hiiumaa. The second clearly differentiated settlements group has weaker social characteristics (no children in villages), than usual for the observable islands. They have small harbours, coastal fishing, summer-cafés, sights etc. There are also no private enterprises officially.

III group, 2003 (2037) < weight, 700 < weight, is main group in both islands.

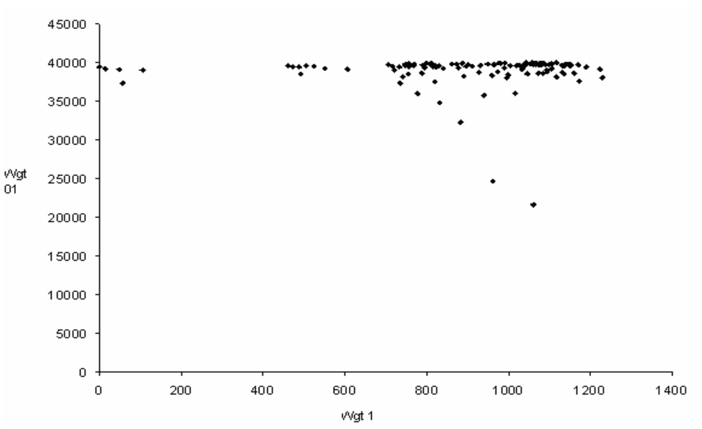
The main group has villages with different social characteristics: some of them have few children, but it contains also settlements that have rank, headmost population-wise for Saaremaa and Hiiumaa. Most of them have some economic indicators.

2.2.3 Composite view from both weight functions

As the both weight functions (π_{01} and π_1) brought forth different aspects, are combining weights of both functions into a single scatterplot (Figure 3.) gave us good overview about settlement groups or clusters, which can be divided into subgroups.



a. Saaremaa



b. Hiiumaa

Figure 3.Conformity plot. Settlements weights by functions π_{01} and $\pi_{1.}$

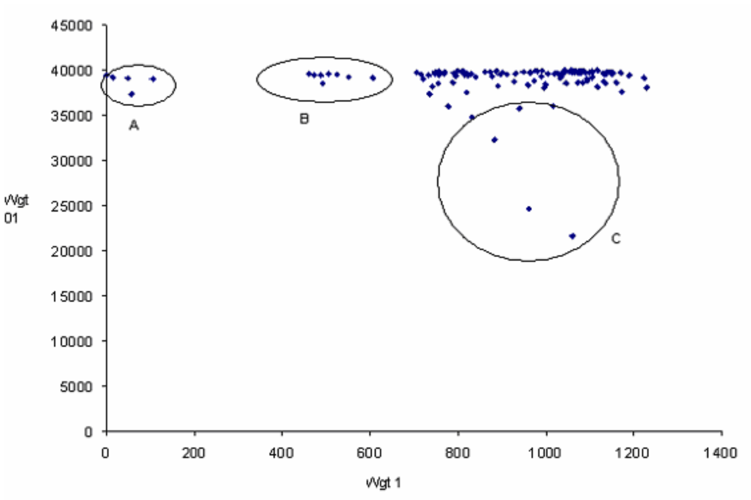


Figure 4. Groups of outliers. (Hiiumaa)

Some easy-to-detect outlier groups are (figure 4):
A: The most non-typical villages, people do not live there. Group I from section 2.2.2.
B: No children in these villages. Group II from section 2.2.2.
C: Large settlements and administrative centers mentioned in section 2.2.1.

Main group can divided into halves or two subgroups.
The first subgroup has villages with different social characteristics: some of them have few children, but it contains also settlements that have rank, third and fourth population-wise for Hiiumaa and fifth, sixth and seventh for Saaremaa and Hiiumaa. Most of them have some economic indicators.
Centre till the end of the main group is formed by most important settlements in islands, included administrative centres. Parts of them are related by strong economic and private investment variables, others are connected by similar social characteristics, like the most typical settlements according to scale of π_{01}-function.
End of the main group represents the most typical villages in scale of π_1-function with more of activities. This group is based on the most widespread social characteristics and activities, like leisure, coastal fishing, small harbour etc. Population of these settlements is 10 till 50 inhabitant, they all have children, workers and elderly.
Structure and semantics of the main group are harder to analyse. Combining conformity plot with information from frequent itemsets or association rules is one promising way to provide semantic information about visual clusters. For example frequent itemset containing villages with workers, elderly and 1 to 10 children splits main groups into halves in both islands.

Hiiumaa`s settlement system is adequately characterized in the map picture, in figure 5.

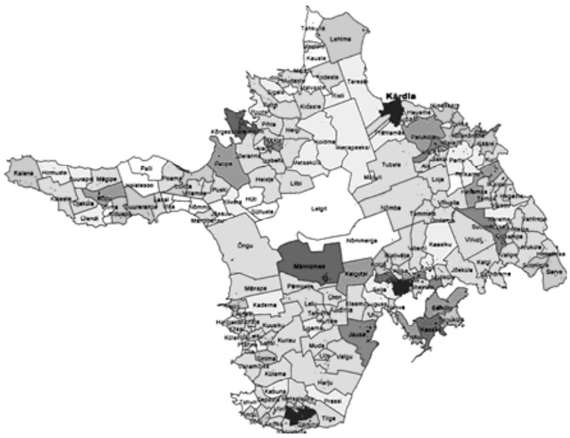


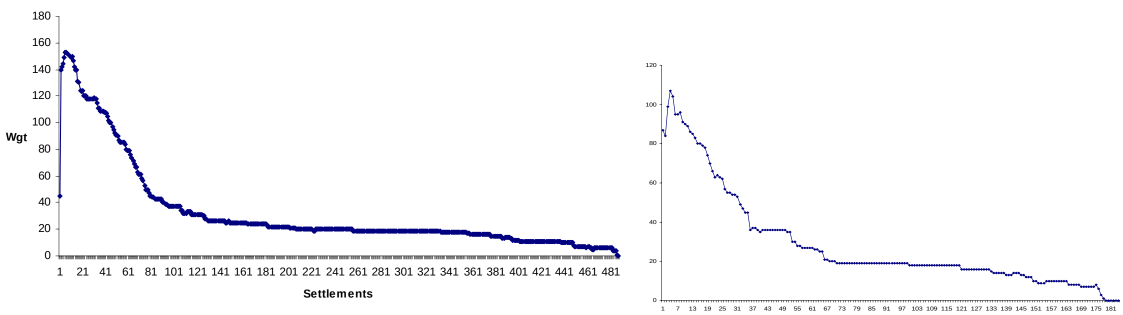
Figure 5. Hiiumaa´s settlements pattern by the scale of conformity. Non-typical centres and developing areas are darker than others (group C). Non-typical villages, where people do not live (group A) are white. Main group (cluster) have represented with gray-scale.

As we can see the graphic structure in figure 3 (a.,b.), 4, bigger settlements have been separated because of the multitude of activities. If the leading settlements have already realized their position in the hierarchy, than recently separated villages have still enough potential to develop, as we can see data analysing with the leading functions.

2.3 Data analysis using the leading functions

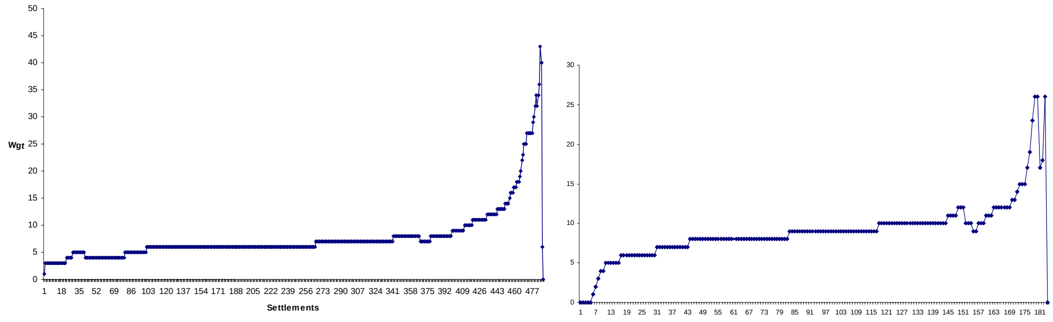
Two different leading functions based also on the conformance scale, but elements or settlements can be eliminate from the table according to their weights, in addition to use algorithms of leading functions, called administrative activeness and incompetence (Kuznetsov, Muchnik, 1982).

The essential difference between these two functions is following:
(1) Administrative activeness is measured by weight function, where the weight is assigned according to such functions that are not represented in other settlements.
(2) Incompetence function assigns the weight to the settlement by subtracting the number of functions that are performed by the settlement from the overall functions that are performed by settlements in general.



a. Saaremaa b. Hiiumaa

Figure 6 Incompetence function. Foremost detach villages with less incompetence.



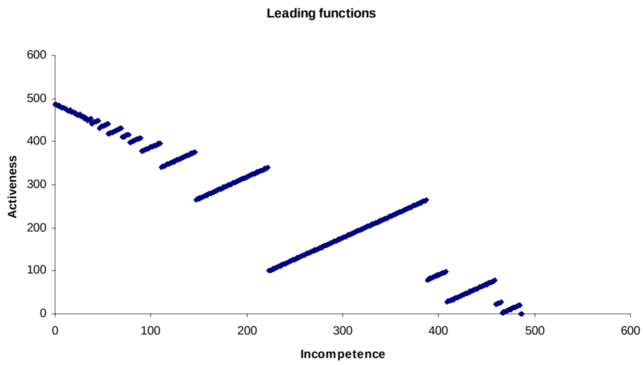
a. Saaremaa b. Hiiumaa

Figure 7 Administrative activeness. Foremost detach villages with less administrative activeness.

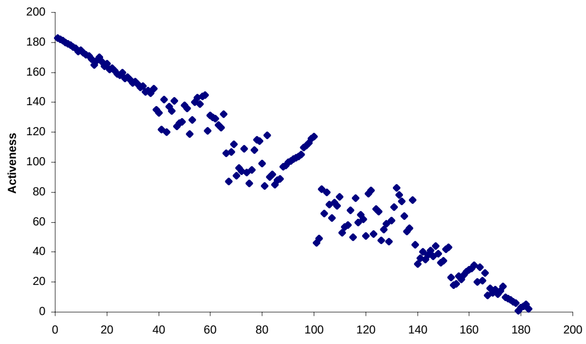
Saaremaa has clear settlements hierachy, as we can see in figures 6a. and 7a.

In Hiiumaa took place overlay about the leading settlements, in the figure 7b. Some villages are as leaders as actors in same time. They are so-called „grey eminences“ (potential new leaders), which have implicated encompassing villages.

Using the algorithms of leading functions are enable to separate all active places and make hierarchical system for the settlements.
If the both leading functions are combining their settlements weights into a single scatterplot (Figure 8 a., b.), than viable villages pull together and push weaker ones to the outside. They form some another groups. In the plot visualization is clear difference between viable active settlements and others with low social characteristics



a. Saaremaa



b. Hiiumaa

Figure 8. Activeness and incompetence scale of the Saaremaa´s and Hiiumaa´s settlements

In the figures 8a and 8b have villages clearly divided into halves: upper ones pull to the activeness, others to the incompetence. It shows us development rank of islands villages. Developing villages are located all in the active half on the scale, these ones and settlements between of the scale own purport to evolve.
The existent leading settlements have already realized their position in the hierarchy, than recently separated villages have still enough potential to develop.

Villages near to the top can be evolved, because of good and convenient characteristics, which support development and are similar to the top villages.

Saaremaa´s and Hiiumaa´s settlements pattern by the activeness as a map picture is brought in figure 9.

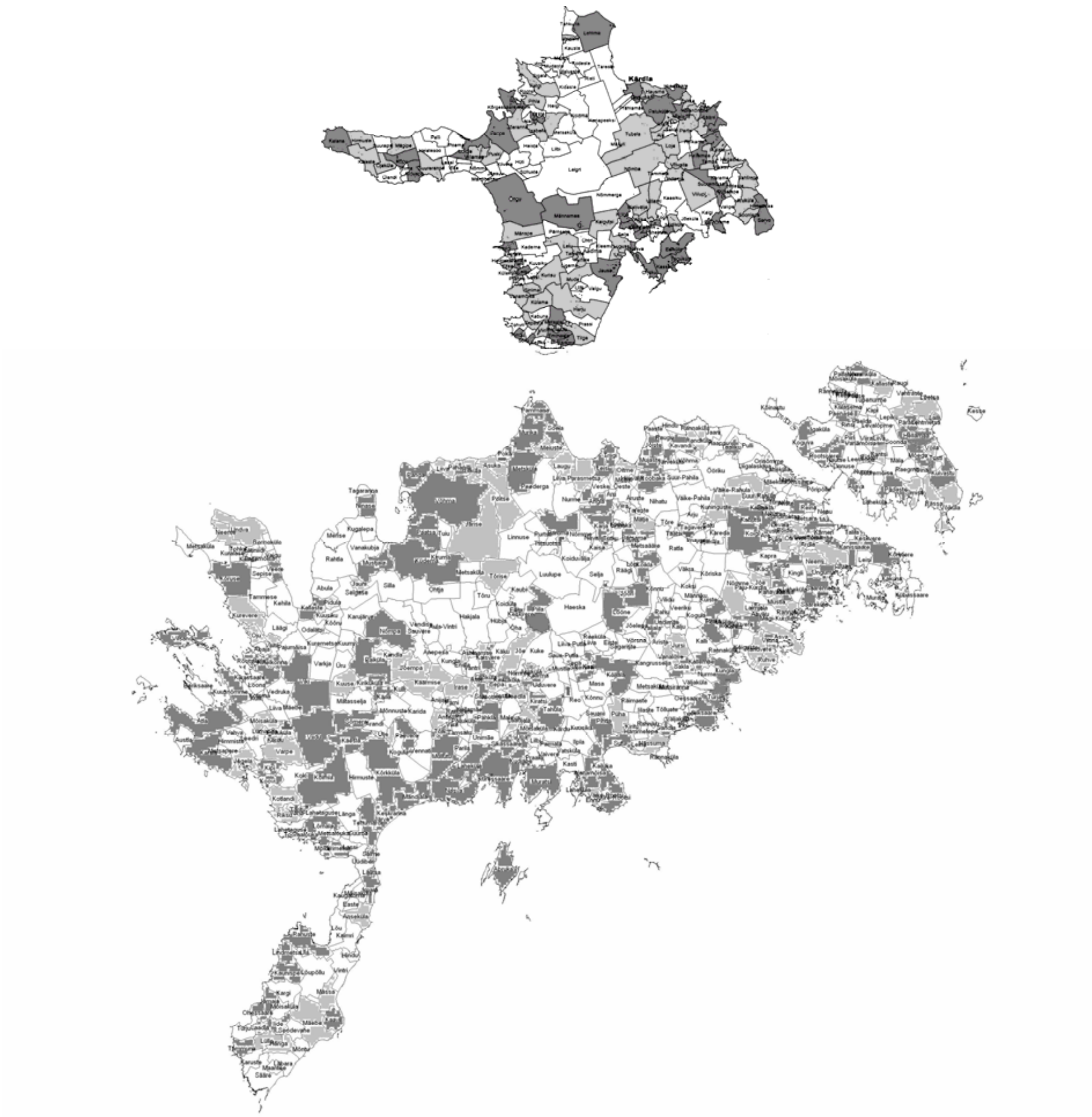


Figure 9 Saaremaa´s and Hiiumaa´s settlements pattern by the activeness.

Centres and developing areas are darker than others. Villages between activeness and incompetence have represented with gray-scale. Incompetent villages are white.

3. Key factors

Hierarchical position of the settlement system is determinated by max 27 (23 in Saaremaa) variables from 254 or 10-12% in Estonian islands (table 3). The most essential are demographic data, like number of people, children etc. Significant are also leisure activities, summerhouses, port or dock, coastal fishing, beach, hostel or camping, shop, sights. Economical affairs have not any importance, exapt farming and turnover of the settlements: in Saaremaa it must be over 1 million Estonian crowns, in Hiiumaa over half million.

Table 3. Most important social-economic characteristics, which have been assigned settlement system in Estonian islands

Saaremaa (between 5-95%)

Characteristic	Code	1- values	Percentage	Comment
library	T8B	26	0,053279	
port	T4	27	0,055328	
people over 100	RA4	32	0,065574	
employees 50-100	TE3	36	0,07377	
shop	T3	39	0,079918	
beach	T15	40	0,081967	
sights	T16	41	0,084016	
Turnover k5	T200	43	0,088115	
farming	T7	55	0,112705	
accommodation	T6A	56	0,114754	
dock	T25	68	0,139344	
people 50-100	RA3	70	0,143443	
children10-50	L2	72	0,147541	
retired10-50	P2	79	0,161885	
summerhouses	T23	87	0,178279	
people10	RA1	104	0,213115	
employees10	TE1	162	0,331967	
employees 10-50	TE2	265	0,543033	(0, 1) weak influence
people10-50	RA2	280	0,57377	(0, 1) weak influence
children10	L1	294	0,602459	
retired10	P1	362	0,741803	
children bt.	L0	377	0,772541	
retired pbt.	P0	449	0,920082	

Hiiumaa (between 5-95%)

Characteristic	Code	1- values	Percentage	Comment
port	T4	10	0,054348	
Turnover k4	T202	10	0,054348	
people over 100	RA4	11	0,059783	
timber or lumber industry	T14B	11	0,059783	
art, music	T29	11	0,059783	
catering (summer)	T6C	12	0,065217	

shop	T3	14	0,076087	
nature protection	T22	14	0,076087	
coastal fishing	T42	16	0,086957	
people50-100	RA3	21	0,11413	
farming	T7	22	0,119565	
beach	T15	22	0,119565	
retired10-50	P2	24	0,130435	
children10-50	L2	27	0,146739	
sights	T16	28	0,152174	
accommodation	T6A	31	0,168478	
people10	RA1	45	0,244565	
summerhouses	T23	62	0,336957	
employeesl10	TE1	73	0,396739	
employees 10-50	TE2	84	0,456522	(0, 1) weak influence
people10-50	RA2	96	0,521739	(0, 1) weak influence
children10	L1	106	0,576087	(0, 1) weak influence
retired10	P1	129	0,701087	
children bt.	L0	140	0,76087	
retired pbt.	P0	157	0,853261	
employees bt.	TE0	171	0,929348	
people bt	T1	174	0,945652	

In the figure 10 we can see the graphic distribution of characteristics.

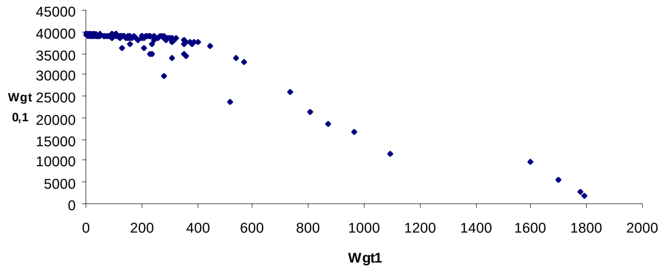


Figure 10. Distribution of characteristics.

All of the outliers have brought in the table 3. Both clearly differentiated point groups have formed by social and demographic data, which are the most important social-economic characteristics in settlements of Estonian islands.

4. Conclusion

Application of monotone systems theory for analysis of social-economic data was successful. We were able to find and describe typical settlements and notable outliers. According to scale of conformance, the outliers are existing top settlements and settlements with lower social characteristics. Leading functions can clearly differentiate the outliers as potential new leaders (central places), fast developing territories and also backward places.

We are using several stages of analysis, but every new stage gave something furthermore to the previous outcomes and supported them. After data processing we can make clear graphic visualization and map pictures to deliver the results. Graphic view afford us clearly differentiated point clusters, which aggregate elements (here settlements) with comparable values. As similar elements have similar weights, it is possible to find groups of similar settlements forming social-economical classification of the settlements.
Same way as the settlements can we analyse the given characteristics. In this case, settlement system can be determinate by less than 12% of social-economic characteristics. Essential are demographic data, but rather is important age-specific structure, than number of population. Significant are also leisure-related activities and coastal fishing, shop, sights, like it was estimated in coastal areas. Economical affairs have not any importance, exapt farming and turnover of the settlements (enterprises).
The main result of the work was presentation of new effective analysis method for regional economics and economic geography. We have collected information about Estonian islands, Saaremaa and Hiiumaa. These both are functional regions, with independent (a little isolated) social life and economy. Target areas are very close neighbours islands and also related historically. That is the reason, why comparison of the results gave use commensurate
Our proposed conformity plot visualization is applicable not only to social data but to all discrete data tables. Also our approach should be practical for analysis of very large data tables.

References

Hoffman, P. E. and Grinstein, G. G., 2002. A Survey of Visualizations for High-Dimensional Data Mining. In Information Visualization in Data Mining and Knowledge Discovery. Academic Press, pp 47-82

Juurikas, K.; Torim, A.; Võhandu, L. 2006. Multivariate Data Visualization in Social Space IADIS International Conference Applied Computing; San Sebastian, Spain; 2006, Feb 25-28, Proceedings of the IADIS International Conference Applied Computing 2006, IADIS Press, pp. 427 - 432

Juurikas,K. and Torim, A., 2006. Data table. http://staff.ttu.ee/~torim/Hiiumaa.xls; http://staff.ttu.ee/~torim/Hiiumaa.csv

Koperski, K. et al, 1998 Mining Knowledge in Geographical Data. In Communications of ACM.

Kuusik, R. and Lind, G, 2003. An Approach of Data Mining Using Monotone Systems. Proceedings of the Fifth International Conference on Enterprise Information Systems. Angers, France. Vol. 2, pp. 482-485.

Kuusik, R. and Lind, 2004. G. New frequency pattern algorithm for data mining. Proceedings of the 13th Turkish Symposium on Artificial Intelligence and Neural Networks. Foca, Izmir, Turkey, ISBN 975-441-213-8, pp. 47-54.

E.N. Kuznetsov, I.B. Muchnik, 1982. Moscow. Analysis of the Distribution Functions in an Organization, Automation and Remote Control, Plenum Publishing Corporation, 1982, pp. 1325-1332,

Mullat, I., 1976. Extremal Monotonic Systems. Automation and Remote Control No 5

Võhandu, L., 1989. Fast Methods in Exploratory Data Analysis. In Transactions of TTU, No 705, pp. 3-13

Võhandu, L. amd Võhandu, P., 2003. Simple and effective methods of data handling in risk analysis. Risk and Safety Management in Industry, Logistics, Transport and Military Service: New Solutions for the 21st Century. Proceedings of the international scientific-educational conference. US Office of Naval Research International Field Office. Technical University of Tallinn. Tallinn, Estonia. Pp. 37-40.