

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/283150158>

Selection of SOM Parameters for the Needs of Clusterisation of Data Obtained by Interferometric Methods

Conference Paper · June 2015

DOI: 10.1109/IRS.2015.7226268

CITATIONS

24

READS

212

2 authors:



Marta Włodarczyk-Sielicka
Maritime University of Szczecin

31 PUBLICATIONS 339 CITATIONS

[SEE PROFILE](#)



Andrzej Stateczny
Gdansk University of Technology

99 PUBLICATIONS 1,211 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Baltic Geodetic Congress (Geomatics) [View project](#)



Sensors Special Issue - Security and Information Flow in Intelligent Systems for the Internet of Things [View project](#)

Selection of SOM Parameters for the Needs of Clusterization of Data Obtained by Interferometric Methods

Marta Włodarczyk-Sielicka^{*}, Andrzej Stateczny^{**}

^{*} Institute of Geoinformatics, Maritime University Szczecin
Waly Chrobrego 1-2, 70-500 Szczecin, POLAND
Email: m.wlodarczyk@am.szczecin.pl

^{**} Marine Technology Ltd.
Klonowica 37/5, 71-248 Szczecin, POLAND
Email: a.stateczny@marinetechnology.pl

Abstract: *The article presents a detailed analysis of parameter settings of self-organizing map (SOM) for the clusterization of bathymetric data obtained using interferometric techniques. Clusterization using SOM is one of the stages of a new geodata reduction method being currently researched by the authors for the purpose of a bathymetric map construction. In the research authors used data obtained by GeoSwath+.-interferometric sonar system. Test data gathered from the area of 100m² included 3760 data points. During the tests the authors focused primarily on setting individual network parameters in the course of network training and also on their importance in the clustering of bathymetric data. A total of forty-eight different scenarios of SOM parameter settings were tested. In the article detailed analysis of the obtained results is presented with an emphasis on the use of SOM in future studies related to new geodata reduction method.*

1. Introduction

The aim of this study is to select self-organizing map (SOM) parameters for the needs of clusterization of data obtained by interferometric methods. SOM is part of a large group of techniques known as artificial neural network (ANN). ANN is very often used in the process of navigation data processing [1-11]. SOM is also called the Kohonen network – an unsupervised neural network algorithm developed by Teuvo Kohonen. In the process of the network's training there is no association between input signals and output of the network. A competition between neurons is a philosophy for the value of its weights updating. During the competition the neuron closest to the input sources in the meaning of chosen distance method calculation is a winner for input data set. Self-organizing maps learn to cluster data based on similarity, topology, with a preference (but not guaranteed) of assigning the same number of instances to each class.

Authors of the hereby article used the high frequency interferometric sonar system GeoSwath Plus for collecting bathymetric data. Bathymetric data is often provided by electronic navigation chart (ENC) producers. Some aspects of electronic charts for inland and maritime navigation, chart production planning and navigational data evaluation are presented in [12-20]. A motor vessel Hydrograf XXI, with the GeoSwath sonar and supplementary equipment (GPS/RTK, satellite compass, motion sensor) installed, was used to collect the data. The GeoSwath sonar uses a different technique of forming an acoustic beam than it is traditionally implemented in multi-beam systems. The measurement is performed by means of a phasic comparison of the measured angle at which the beam is deflected off the bottom. Each of the two heads of the system is comprised of a single transmitting disc and multiple receiving discs of similarly shaped beams. The amplitude and phase of the signal dispersed off the bottom surface are measured by electronic circuitry. The relative phasic delay (time delay measurement) between the head's receiving discs is decoded as the angle of the acoustic

signal's return. The measurement accuracy of the return signal's angle is measured to a fraction of a degree at very short time intervals, resulting in the high resolution of the data. The beam formed out of transducers set at a 30° angle has similar properties as the one generated in sonars – the angle width along the beam is determined above 150°, the width across – about 1°[21]. Contemporary bathymetric measurement systems allow to record more and more data of increasingly high quality. Interferometric systems represent systems able to collect high density data.

2. The specification of geodata reduction method

The main aim of authors is to create new reduction method for high-density bathymetric data in terms of the construction of the bathymetric map. Spatial clustering means putting similar features into the same group. Clusterization using SOM is one of the stages of new geodata reduction method. The final hydrographic product, which is the bathymetric map, should be legible, while presented data should comply with the relevant guidelines. Suitable gathering, preparation and presentation of data lies with a hydrograph. Processing of the data is a long and laborious process. In order to correctly execute all the stages, a new algorithm should be created that would allow, among other things, the removal of unnecessary observations (observations removed without affecting the quality of the bathymetric map visualization), and, consequently, would reduce the time and costs of data processing.

Data obtained using the interferometric system consists of a very large set of measurement points and there is a necessity to reduce them. For example, in the area of 625m² there were received 28911 measurement points. To make the reporting site plan at the scale of 1:1000 legible, data needs to be reduced to about 312 points. Data reduction is a procedure of reducing the size of the data set in order to make it easier and more efficient for analysis. In addition to the accuracy and reliability of the obtained results, the dynamics of information processing and the ability to perform rapid analysis is essential as well.

Software used to process bathymetric data usually base on creating GRID nets, in which nodes are interpolated values. Generally, the hydrographic systems generate grids with the use of one of the following methods: mean (select a mean depth value) or weighted mean (uses amplitude values to give higher weighting to data points which are higher in amplitude when calculating the mean depth value). There is therefore a need to automatically present depths on the bathymetric map, regardless of the scale of its development, while maintaining the required fidelity of the surface mapping and while, at the same time, maintaining positions and values of minimum depths which are essential for the safety of navigation. Efficient use of large data sets is a problem in the processing, analysis, and sharing of data.

One of the stages of the new method being developed is data clustering using artificial neural networks. It is based on the SOM map, also known as Kohonen neural network. This network is based on learning "without a teacher", which means that there is no need to intervene during the learning. The network is trying to learn the data structure on its own. Neurons presented on the map are connected to adjacent neurons by neighborhood relationships – that's how the structure (topology) of the network is created. The network takes into account not only the X and Y coordinates of data points, but also a depth value which is assigned to each of them.

KSelf-organizing maps learn to cluster data based on similarity, topology, with a preference (but not guaranteed) of assigning the same number of instances to each class. In Matlab software SOM inputs such arguments as: layer topology function, row vector of dimension sizes, number of training steps for initial covering of the input space, initial neighborhood size and neuron distance function.

3. The experiments

All test methods are implemented using Matlab software developed by MathWorks. To collect bathymetric data the vessel Hydrograf XXI, with the GeoSwath Plus 250 kHz sonar

and supplementary equipment like GPS/RTK, satellite compass and motion sensor installed was used.

During the studies, test data gathered from the area of 100m² was used, which includes 3760 data points, as shown on Figure 1. The minimum depth of the test area is 3.6 meters and the maximum – 5.23 meters. The positions of data are given by the Universal Transverse Mercator coordinate system, which is an international locational reference system. Each point has three attributes: latitude, longitude, and a predetermined depth at a given point.

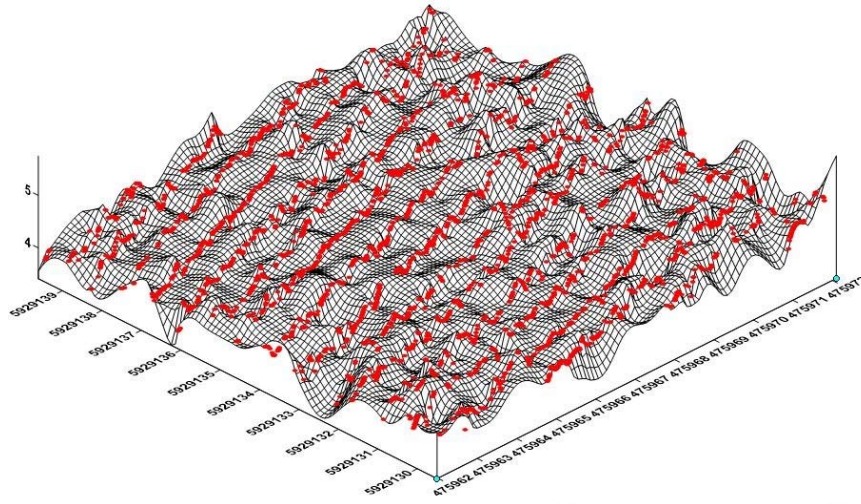


Figure 1. Test data on the area 100 square meters [own studies]

During the use of Kohonen's self-organizing maps, the network, during learning, use the WTM rule - rule, which modifies not only the weight of the winner, but also its neighbors. The hexagonal topology is selected (each of the hexagons represents a neuron) and the number of epochs is set at 200 based on recent research [22]. The numbers of rows and columns are set to 5x5 and 3x3, which provide the number of neurons 25 and 9. Number of training steps for initial covering of the input space is set at 100. Distances from a particular neuron to its neighbors are implemented with four different function:

- Euclidean distance - the straight line distance between two points;
- link distance - the link distance from one neuron is the number of links or steps that must be taken to get to the neuron under consideration;
- box distance - the layer distance function that is used to find the distances between the layer's neurons, given their positions;
- Manhattan distance - the distance between two points measured along axes at right angles.

Initial neighborhood size parameter is set at 1, 3, 5, 10, 100 and 1000. The total yield of 48 different sets of clusters are analyzed.

4. The results

The results are examined depending on the settings of four different distance functions and six different parameter settings for initial neighborhood size. During the research twenty-four scenarios for nine clusters (received two hundred sixteen individual clusters) and twenty-four scenarios for twenty-five clusters (received six hundred individual clusters) were analyzed. During the analysis some statistics for the depth information were automatically calculated, because of its key value as bathymetric data and its significant importance for the safety of navigation. They were: minimum value of depth, maximum value of depth, mean value of depth, number of samples in clusters and standard error of the mean (SEM) for each cluster. The authors assumed that small value of SEM is closely related to regular distribution of

bathymetric data in each cluster. SEM is the standard deviation of the sampling distribution of the mean. SEM is estimated by the sample estimate of the population standard deviation divided by the square root of the sample size.

The following criteria of evaluation are accepted: ability to automate, time of calculations and data distribution in each cluster. In the next step the comparison analysis of numbers of samples in single clusters for each scenario was made. It can be assumed that in the case of 9 clusters the average number of points in each cluster is approximately 418 samples. This comparison is shown in Figure 2.

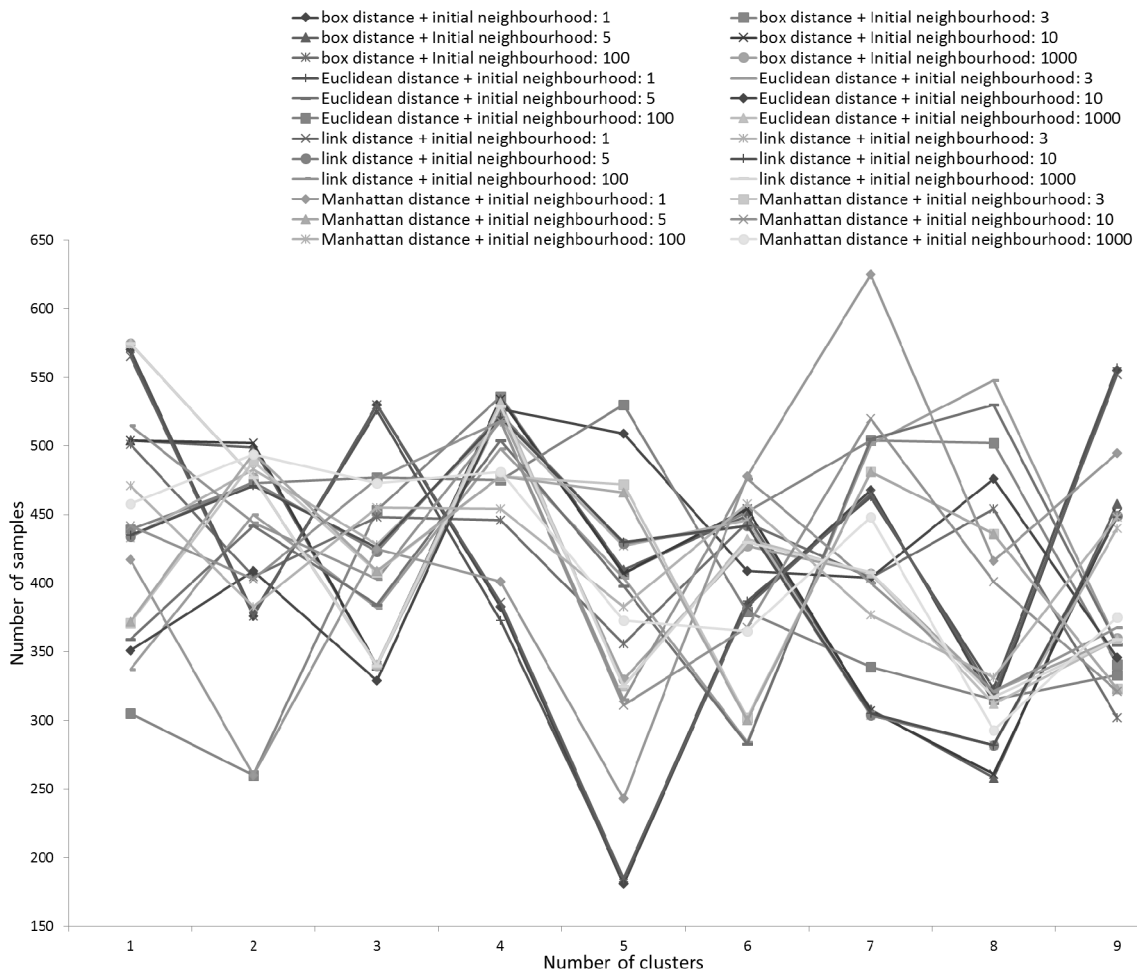


Figure 2. Distribution of the number of samples in each cluster.

The horizontal axis represents the number of clusters and the vertical axis shows number of bathymetric data points. This comparison visualizes a distribution of data in each cluster for several scenarios. Each of the established criteria is considered in a relation to the use of clusterization for very-high density bathymetric data. After a detailed analysis of all the results, taking into account the established criteria, the following settings for SOM were selected: Manhattan distance function and initial neighborhood size parameter set at 100. Fig. 3 shows spatial representation of the final results for the selected settings of SOM for 9 clusters. Fig. 4 presents numbers of samples in clusters for the selected settings of SOM for 9 clusters. The axis X represents the number of clusters and the axis Y shows number of samples. Because the focus lied on depth values the statistical analysis was conducted mainly in relation to this information of the sample. Tab. 1 presents the statistic results for the selected settings of SOM for 9 clusters.

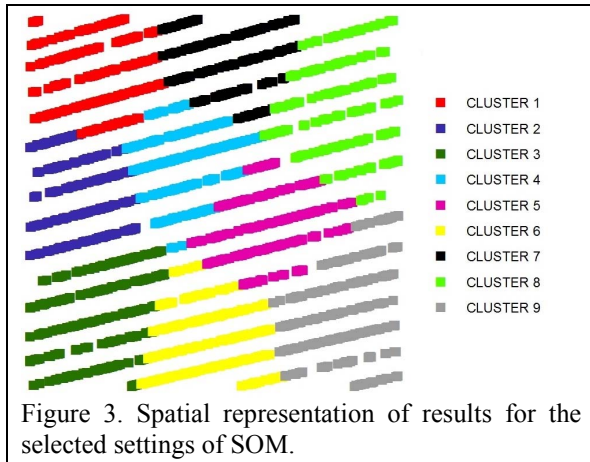


Figure 3. Spatial representation of results for the selected settings of SOM.

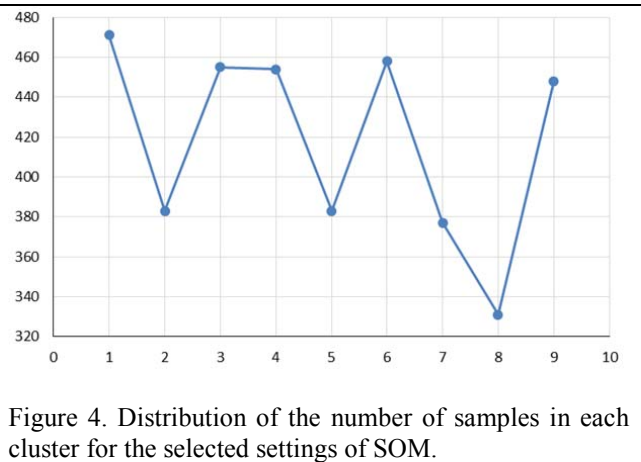


Figure 4. Distribution of the number of samples in each cluster for the selected settings of SOM.

Table 1. Statistic results for the selected settings of SOM

Cluster	Minimum of depth [m]	Maximum of depth [m]	Mean of depth [m]	Standard deviation	Number of objects in cluster
1	3,7	4,6	4,181953	0,168242	471
2	3,66	4,61	4,125901	0,17771	383
3	3,6	4,74	4,230615	0,18408	455
4	3,98	4,89	4,405	0,165448	454
5	4,03	5,23	4,59329	0,220424	383
6	3,99	4,95	4,408057	0,168136	458
7	3,93	5,2	4,547745	0,189535	377
8	4,09	5,22	4,640755	0,230817	331
9	4,16	5,2	4,715804	0,199593	448

The initial neighborhood size parameter is of a great significance. It should be remembered that application of different test scenarios results in different nomenclature of clusters. However, this is not important in reduction of bathymetric data.

5. Conclusions

The main purpose of authors research is to create a new data reduction method. Clusterization using SOM is one of the stages of this method. In the analysis of the test results the authors focused on depth information carried by third dimension of a sample. Analyzing the tests performance it can be noticed that the best results are received using the following settings of SOM: WTM rule, hexagonal topology, number of epochs set at 200, number of training steps for initial covering of the input space set at 100, Manhattan distance function and initial neighborhood size parameter set at 100. These settings of SOM will be used for future research on reduction algorithm for bathymetric data.

References:

- [1] J. Balicki, Z., Kitowski, A. Stateczny, "Extended Hopfield Model of Neural Networks for Combinatorial Multiobjective Optimization Problems", *Proc. of 12th IEEE World Congress on Computational Intelligence*, Anchorage, USA, 1998, pp. 1646-1651.
- [2] W. Kazimierski, A. Stateczny, "Optimization of multiple model neural tracking filter for marine targets", *Proc. of International Radar Symposium*, Warsaw, Poland, 2012, pp. 543-548.
- [3] W. Kazimierski, G. Zaniewicz, A. Stateczny, "Verification of multiple model neural tracking filter with ship's radar", *Proc. of International Radar Symposium*, Warsaw, Poland, 2012, pp. 549-553.

- [4] J. Lubczonek, A. Stateczny, "Concept of neural model of the sea bottom surface", *Proc. of Neural Networks and Soft Computing*, Zakopane, Poland, 2003, *Advances in Soft Computing*, pp. 861-866.
- [5] A. Stateczny, W. Kazimierski, "A comparison of the target tracking in marine navigational radars by means of GRNN filter and numerical filter", *Proc. of IEEE Radar Conference*, Rome, Italy, 2008, vols. 1-4, pp. 1994-1997.
- [6] A. Stateczny, W. Kazimierski, "Determining Manoeuvre Detection Threshold of GRNN Filter in the Process of Tracking in Marine Navigational Radars", *Proc. of International Radar Symposium*, Wroclaw, Poland, 2008, pp. 242-245.
- [7] A. Stateczny, W. Kazimierski, "Selection of GRNN network parameters for the needs of state vector estimation of manoeuvring target in ARPA devices", *Proc. of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, Wilga, Poland, 2006, vol. 6159, pp. F1591-F1591.
- [8] A. Stateczny, "Artificial neural networks for comparative navigation", *Proc. of Artificial Intelligence and Soft Computing*, Zakopane, Poland, 2004, *Lecture Notes in Artificial Intelligence*, vol. 3070, pp. 1187-1192.
- [9] A. Stateczny, "Neural manoeuvre detection of the tracked target in ARPA systems", *Proc. of Control Applications in Marine Systems*, Glasgow, Scotland, 2001, *IFAC Proceedings Series* 2002, pp. 209-214.
- [10] A. Stateczny, "Methods of comparative plotting of the ship's position", *Proc. of Maritime Engineering & Ports III*. Rhodes, Greece, 2002, *Water Studies Series*, vol. 12, pp. 61-68.
- [11] A. Stateczny, "The neural method of sea bottom shape modelling for the spatial maritime information system". *Proc. of Maritime Engineering and Ports II*, Barcelona, Spain, 2000, *Water Studies Series*, vol. 9, pp. 251-259.
- [12] W. Kazimierski, A. Stateczny, "Fusion of Data from AIS and Tracking Radar for the Needs of ECDIS", *Proc. of Signal Processing Symposium*, Jachranka, Poland, 2013.
- [13] W. Kazimierski, A. Stateczny, "Radar and Automatic Identification System track fusion in an Electronic Chart Display and Information System", *Journal of Navigation* 2015 (in press).
- [14] A. Stateczny, W. Kazimierski, "Sensor Data Fusion in Inland Navigation", *Proc. of 14th International Radar Symposium*, Dresden, Germany, 2013, vols. 1 and 2, pp. 264-269.
- [15] A. Stateczny A., I. Bodus-Olkowska, "Hierarchical Hydrographic Data Fusion for Precise Port Electronic Navigational Chart Production", *Proc. of Telematics in the Transport Environment*, Ustron, Poland, 2014, *Communications in Computer and Information Science* 471, pp. 359-368.
- [16] J. Lubczonek, A. Stateczny, "Aspects of spatial planning of radar sensor network for inland waterways surveillance", *Proc. of 6th European Radar Conference*, Rome, Italy, 2009, pp. 501-504.
- [17] A. Stateczny, J. Lubczonek, "Radar Sensors Implementation in River Information Services in Poland", *Proc. of 15th International Radar Symposium*, Gdansk, Poland, 2014, pp.199-203.
- [18] A. Stateczny, J. Lubczonek, "FMCW radar implementation in River Information Services in Poland", *Proc. of 16th International Radar Symposium*, Dresden, Germany, 2015.
- [19] A. Stateczny, I. Bodus-Olkowska, "Sensor data fusion technics for environment modelling", *Proc. of 16th International Radar Symposium*, Dresden, Germany, 2015.
- [20] A. Stateczny, J. Lubczonek, T. Kantak, "Radar Sensors Planning for the Needs of Extension of River Information Services in Poland", *Proc. of 16th International Radar Symposium*, Dresden, Germany, 2015.
- [21] Llort-Pujolv G. et al., "Advanced interferometric techniques for high-resolution bathymetry", *Journal of Marine Technology Society*, Volume 46, Number 2, March/April 2012, pp. 9-31
- [22] A. Stateczny, M. Wlodarczyk-Sielicka, "Self-Organizing Artificial Neural Networks into Hydrographic Big Data Reduction Process", *Proc. of Joint Rough Set Symposium*, Granada-Madrit, Spain, 2014, *Lecture Notes in Artificial Intelligence*, pp. 335-342.