

DEVELOPMENT OF AN INTEGRATED ANN-GIS FRAMEWORK FOR INLAND EXCESS WATER MONITORING

Van Leeuwen, B.¹ – Tobak, Z.¹ – Szatmári, J.¹

¹Department of Physical Geography and Geoinformatics, University of Szeged, Hungary

Abstract

Inland excess water on the Great Hungarian plain is an environmental and economic problem that has attracted a lot of scientific attention. Most studies have tried to identify the phenomena that cause inland excess water and combined them using regression functions or other linear statistical analysis. In this article, a different approach using a combination of artificial neural networks (ANN) and geographic information systems (GIS) is proposed. ANNs are particularly suitable for classifying large complex non-linear data sets, while GIS has very strong capabilities for geographic analysis. An integrated framework has been developed at our department that can be used to process inland excess water related data sets and use them for training and simulation with different types of ANNs. At the moment the framework is used with a very high resolution LIDAR digital elevation model, colour infrared digital aerial photographs and in-situ fieldwork measurements. The results of the simulations show that the framework is operational and capable of identifying inland excess water inundations.

Keywords: inland excess water, artificial neural network, geographic information systems

INTRODUCTION

Inland excess water is a reoccurring problem in the Great Hungarian Plain. At the end of winter large parts of the flat terrain are covered by water. These inundations cause serious economic and environmental problems.

Several studies have analysed the problem, with varying success (Bozán Cs. et al. 2005, Pásztor L. et al. 2006, Rakonczai J. et al. 2001, Rakonczai J. et al. 2003). Most studies have tried to identify the phenomena that cause the inland excess water and combined them using regression functions or other linear statistical analysis. In this article, a different approach using artificial neural networks (ANN) is proposed. This approach has many advantages compared to other statistical methods. First, it is independent of the statistical distribution of the data, and there is no need for specific statistical variables. Neural networks allow the target classes to be defined in relation to their distribution in the corresponding domain of each data source, and therefore the integration of remote sensing or GIS data is very convenient (Pradhan B. – Lee S. 2010).

Certain types of inland excess water can be forecast and those areas or points where action is needed for decreasing or even avoiding damage can be directly determined with the help of theoretical and practical means. This way the risk of inundation can be mitigated in numerous occasions, and this could lead to a shift from a reactive, defensive-type water management strategy towards a more proactive strategy, in order to decrease or even prevent damage.

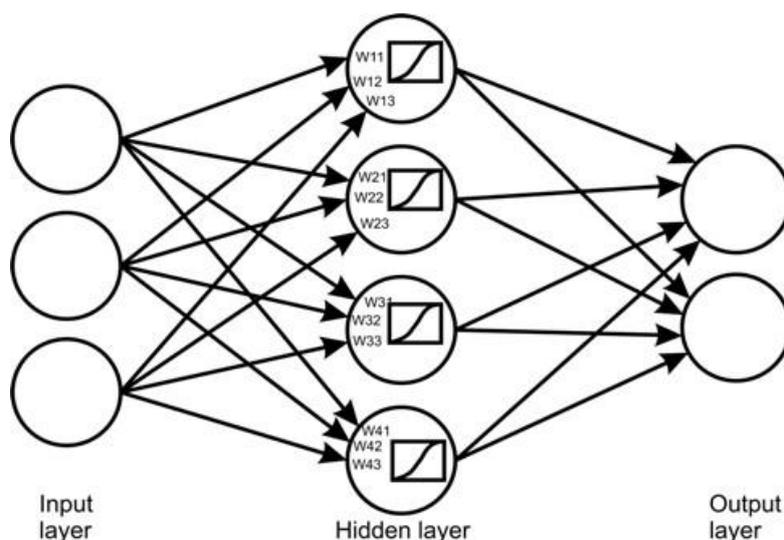


Fig. 1 A basic artificial neural network

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are computational models that imitate the functioning of the human brain. Several different types of neural networks exist but their basic structure always consists of multiple layers of interconnected nodes (*Fig. 1*). Every neuron processes the weighted sum of all inputs, and, via a so-called activation function it is determined if the signal is sent further.

The application of ANNs consists of two phases. The first phase is called the training phase. During this phase the ANN is fed with an input and an associate output data set. The training is an iterative process where the weights of the incoming signals are adapted in such a way that the overall average error between the requested output and the calculated results is minimized. The trained network can be used in the second phase where it is fed with new input data to calculate new output results. A more detailed description of ANNs goes beyond the scope of this article but can be found in Retter Gy. (2006), Hewitson B. C. and Crane R. G. (1994) and Zurada J. M. (1992).

ANNs have been proven themselves in many fields of science where complex data sets need to be analyzed to identify their underlying structures and properties. Neural networks have a large potential for analysis of complex spatial problems which are common in geographic research (Hewitson B. C. – Crane R. G. 1994). Inland excess water inundations on the Great Hungarian Plain are a clear example of such problems. The reoccurring inundations are caused by a multitude of interrelated factors.

The connection between the world of neural networks and GIS is still relatively new and needs to be developed further (Coleman A. 2008, Sárközy F. 1998). Just two GIS software exist that employ fully integrated GIS – neural network solutions; ArcGIS and IDRISI. These solutions have been investigated but were not used in this study because they employ only one type of neural network architecture, a multi-layer perceptron and a radial basis function network, respectively, and they do not offer integrated tools for the evaluation of the training and the simulation results.

Matlab 7.10.0 has an integrated neural network

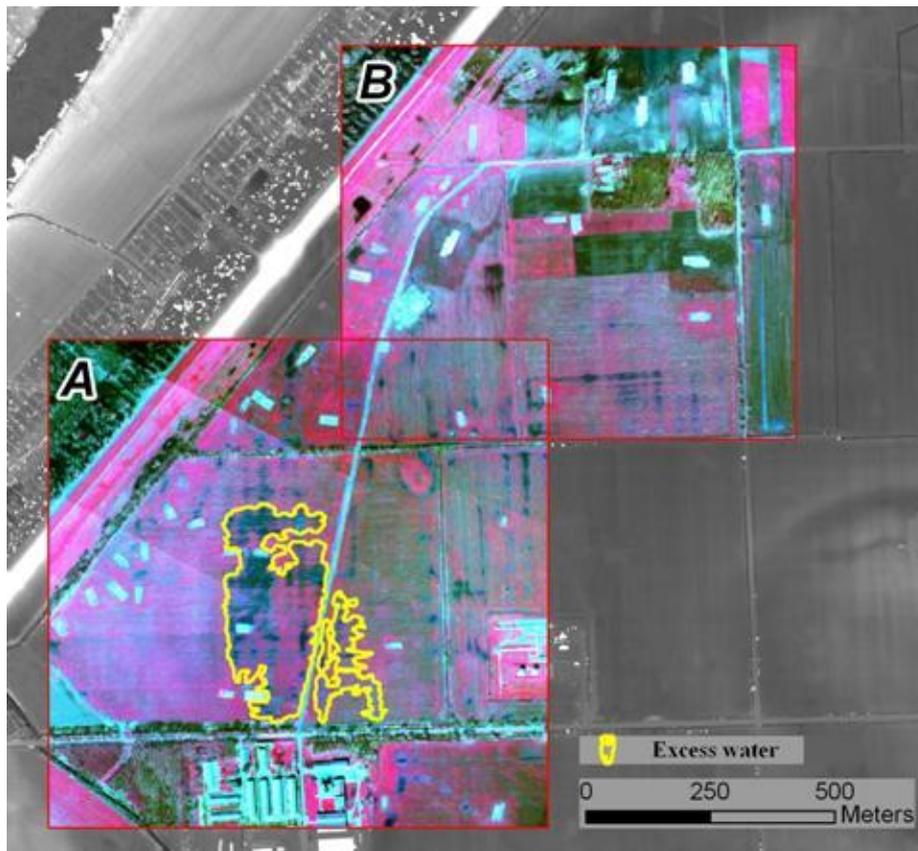


Fig. 2 The digital elevation model overlaid with CIR mosaics of 24 March 2010 showing the training (A) and simulation (B) area with the GPS fieldwork results in the Tápairét area

toolbox that ranges from simple solutions to extended neural network implementations. The determination of the network architecture constitutes one of the major and most difficult tasks in the use of neural networks (Barsi Á. 1997, Jafar R. et al. 2010). Since it is not exactly known what type of neural network with which settings is most appropriate to study the problem of inland excess water, it was decided to build a framework that facilitates the possibilities to experiment with several neural networks and settings in a GIS environment. ArcGIS 9.3 was used as the GIS environment because of its strong capabilities for geographic analysis, and its possibilities for customization.

STUDY AREA AND DATA

The Great Hungarian Plain covers an area of 52,000 km². The Tápairét area was selected from this region as a test site for the inland excess water research (Fig. 2). This study area is about 20 km² large and its maximum difference in elevation is 10 meters. Mainly agricultural activity takes place in the area, although there are also several oil stations. From the young sediments with high clay content of the Maros River, fluvisols and vertisols were formed (Marosi S. – Sárfalvi B. 1990). Because of the extreme mechanical properties – in large areas, the plasticity index (K_A) is above 60 (cm³/gr) –, the exceptionally bad permeability characteristics result in accumulation of water in the lower areas.

Table 1 gives an overview of the data used in this research. All data were collected in the period 2009-2010.

Apart from the bad soil characteristics, the area consists of very flat terrain with large local depressions, without run-off. The average groundwater level varies between 2 and 4 meters below the surface. Remnants of river meanders can also be found in the area. Only in the former meanders, the groundwater may reach the surface. This research focuses on the genetic type of the inland excess water that is caused by a lack of runoff and infiltration, and not on the type that is due to high

groundwater levels.

FRAMEWORK

A framework was created to handle input data, intermediate results and output data in a flexible way in both ArcGIS and Matlab (Fig. 3). In this way, it was possible to create the data files, test different network types and settings and evaluate the training and simulation results efficiently.

First, different artificial data sets were created in ArcGIS. These data sets were used to set up the framework and to evaluate the simulated results. Three artificial input maps of 100 by 100 pixels were created. Each map represented specific inland excess water related input parameters (e.g. local depressions, geomorphologic structures, soil types, height of the groundwater table, land use). A fourth artificial map was created to represent the occurrences of inland excess water in the same area. The files were created using ArcGIS 9.3 and were stored in TIFF file format. The TIFF files were read into Matlab resulting in a 100x100 cell matrix for each map.

The neural network analyses were performed with the neural network toolbox of Matlab. This is an extension of the general Matlab functionality incorporating many artificial neural network architectures and tools for training and evaluation of the results (Demuth H. et al. 2010). The neural network toolbox needs data in a matrix format where every row represents an input data layer. A program was written to convert the separate input matrices to arrays and to combine the resulting 1x10000 arrays into one matrix with 3x10000 cells that could be read by the neural network toolbox. The output matrix, representing the occurrences of inland excess water was converted to an 1x10000 array as well. With the artificial data, only the standard neural net in the nftool from the neural network toolbox was used. This is a two layer feed forward network with maximum 20 neurons in the hidden layer. A smaller amount of neurons gave similar results but resulted in lower performance due to more iterations. The network was trained

Table 1 Input and output data

LIDAR /DEM/ local depressions	LIDAR data with a spatial resolution of 1.4 points per m ² were collected from a 70 km ² area during a flight campaign on 19 November 2009. Based on this data, a 1 meter resolution digital elevation model was created.
CIR (Colour-InfraRed) imagery	At the maximum of the inland excess water periods, on 24 March and 9 June 2010, flights were executed using a data collection system based on the MS3100 digital camera (Tobak Z. et al. 2008) to collect 800x600 meter images. From all individual images a 63 cm resolution mosaic covering an area of 60 km ² was created.
Field measurements	On 5 March, 2010, a one day fieldwork was carried out in the south-western part of the study area. At that moment, the second level on the national inland excess water hazard scale was valid. In total 7.8 ha of inundated land was accurately measured by walking around them using hand-held GPS systems.

with 70% of the data, while 15% was used for validation and 15% for testing. The optimal network was saved to be used in the simulation phase. Simulation data was then imported from the GIS, converted to a matrix and fed to the neural net.

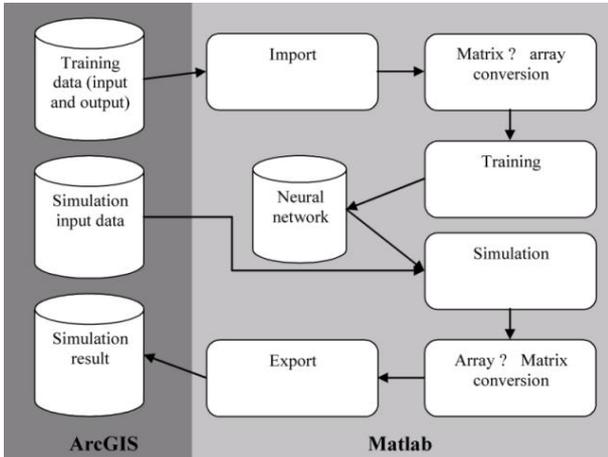


Fig. 3 Framework showing the workflow in ArcGIS and Matlab

The simulation result was again converted into a matrix. During the different conversion steps the data had to undergo various types of conversions to be compliant with the particular data formats. Finally, a continuous 8 bit TIFF file was generated which could be visualized in ArcGIS.

Apart from several pre-processing steps, the same workflow, as described above with the artificial data, was followed using the new, real data set as well (Fig. 4). The training data consisted of 4 input and one output layers. The colour infrared images were split in three bands; green, red and near infrared. Using the fill tool in

ArcGIS, the local sinks in the LIDAR based digital elevation model were filled (Tarboton D. G. et al. 1991). The original height values were subtracted from the sink map, resulting in a layer with the local depressions. The depression map was reclassified into three classes: very small depressions (<15 cm), middle (15-60 cm) and deep (>60 cm) depressions. The resulting map was used as the fourth input layer in the training phase. The fieldwork measurements were rasterized and used as output map during the training. That time only two output classes were defined: open water and dry soil. Every data layer had a spatial resolution of 1 meter and was covering an area of 1000x1000 meter.

During the simulation phase the same type of CIR imagery and elevation data were used. The same pre-processing steps were executed as in the training phase; just the location of the data was several hundreds of meters further to the north (Fig. 2).

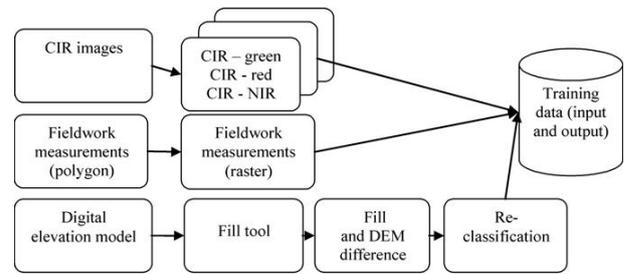


Fig. 4 The pre-processing of the training data

RESULTS

Several settings for the number of neurons in the hidden layer were tested. With an increase of the neurons, the RMSE decreased, however, the performance of

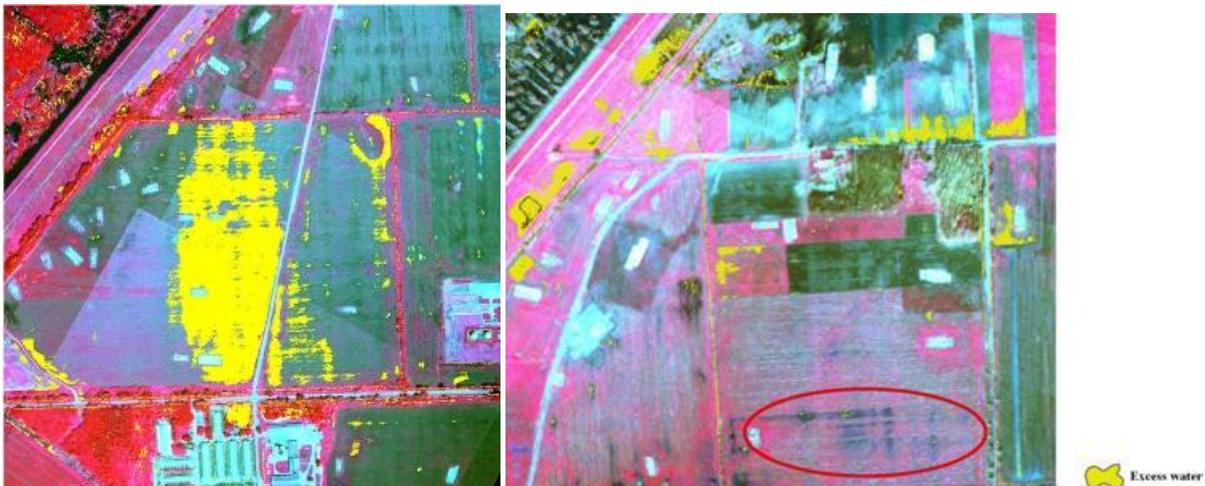


Fig. 5 The results of the training (left) and the simulation (right)

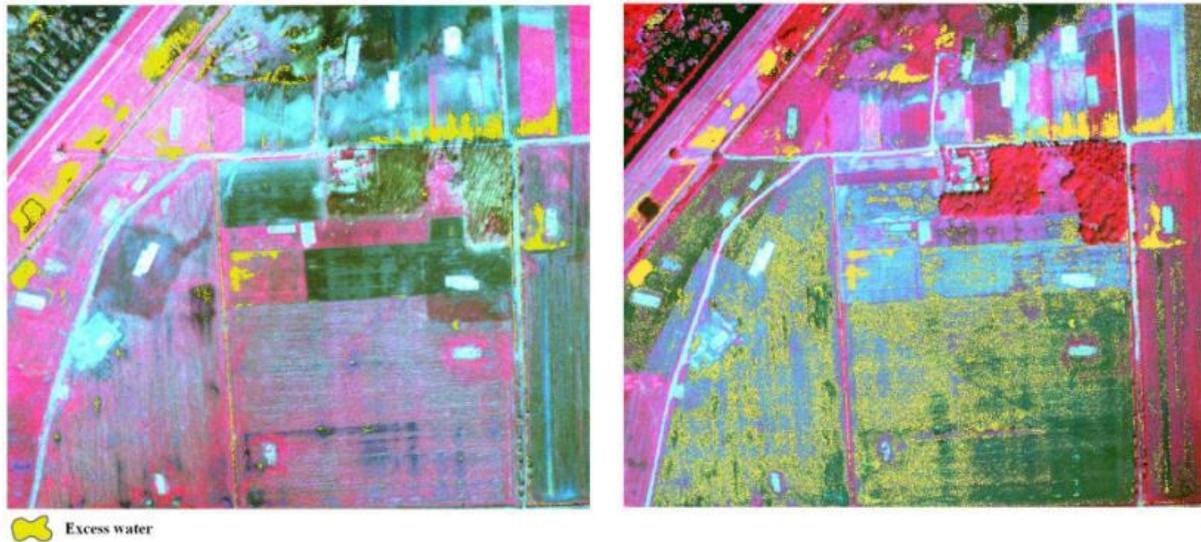


Fig. 6 The simulation results of two different times: 24 March 2010 and 9 June 2010

the training also decreased sharply. An optimum of 10 neurons was selected resulting in an overall RMS training error of 0.74. The result of the training is shown on the left side of *Fig. 5*.

The right side of *Fig. 5* shows the result of the simulation using the trained network. The yellow areas were classified as inland excess water. In the northern and north-western part of the area the results are good. The open water along the levee and the roads was detected. The inland excess water in the southern part of the images is not properly classified. Some pixels are correctly indicated as inland excess water but the majority is classified as dry land. These errors are probably due to the composition of the training set, where only water was incorporated but saturated soil and vegetation in water were omitted.

A second simulation was executed using the same trained ANN, but this time with different multi-spectral data. In this simulation, the colour infrared images collected on 9 June were combined with the same local depression data that was used in the first simulation (*Fig. 6*).

Although in general, the inland excess water areas that were identified on the images taken on 24 March, were also classified as water on the images taken on 9 June, on the second date much more inland excess water was detected. Furthermore, the second simulation shows that there is scattered water on the large parcels in the centre of the images. This may indicate that the soil in this area was completely saturated with water. Since no ground truth was collected for the area at the time of the data acquisition, it is not possible to quantify the simulation differences.



Fig. 7 Comparison of different classification methods: maximum likelihood (left), minimum distance (middle), artificial neural network (right). White colour indicates inland excess waters, all other areas are in black. The training data is shown with a red-coloured boundary

A comparison has been executed among the training results of the ANN and two traditional classification methods: maximum likelihood and minimum distance (Fig. 7). The ANN classification clearly shows the white area overlapping with the training area. Several other patches of inland excess water were also classified. For these areas no ground data was collected but they can easily be identified visually on the CIR images (Fig. 2/A). For the other two classifications only the pixels of the training area were used during the supervised training. For both traditional classifications this results in accurate classification of the inland excess water in the training data, but also in an extreme over-classification in the areas outside this area.

CONCLUSIONS

The framework works as expected with a small artificial test data set. The larger real data set also resulted in proper delineation of inland excess water, but further development is still needed. Due to the nature of spatial data, very large matrices are created as input data for the network. This results in performance problems. By reducing the amount of input pixels in the input data sets, the performance of the system can be improved. The result of the simulation shows a clear distinction between water and dry soils. In reality this is a fuzzy boundary. Intermediate classes like saturated soil and vegetation in water also exist. These classes were not taken into account in the training set. Extra field data will be needed to incorporate these classes and to be able to derive them in the simulation. This fieldwork data is also needed to be able to quantify the differences in results between the different classification methods. Furthermore, other input data sources, like soil maps, hydrological maps can be incorporated to extend the base of the training set. Finally, the integration between the GIS and the neural network has to be improved. The framework now consists of several loosely coupled programs and Matlab functions. To facilitate the most efficient prototyping their integration is inevitable.

Acknowledgement

This study was financially supported by the project "Development of an INLAND EXCESS WATER-INFO system" (Economic Operative Program: GOP - 1.1.1 - 08 / 1 -2008 – 0025).

References

- Barsi Á. 1997. Landsat-felvétel tematikus osztályozása neurális hálózattal. *Geodézia és Kartográfia* 49/4: 21-28
- Bozán Cs. – Pálfi I. – Pásztor L. – Kozák P. – Körösparti J. 2005. Mapping of Inland excess water Hazard in Békés and Csongrád Counties of Hungary. ICID 21st European Regional Conference 2005 – 15-19 May 2005. Frankfurt (Oder) and Slubice – Germany and Poland
- Coleman A. M. 2008. An adaptive Landscape classification procedure using geoinformatics and artificial neural networks. Unpublished MSc thesis, Amsterdam. 195 p
- Demuth H. – Beale M. – Hagan M. 2010. Neural Network Toolbox 6, User's Guide. The Mathworks 901 p
- Hewitson B.C. – Crane R.G. 1994. Neural Nets: Applications in Geography. Dordrecht: Kluwer Academic Publishers. 194 p
- Jafar R. – Shahroui I. – Juran I. 2010. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling* 51: 1170-1180
- Marosi S. – Sárfaivi B. 1990. Magyarország kistájainak katasztere I. Budapest: MTA FKI. 1023 p
- Pásztor L. – Pálfi I. – Bozán Cs. – Körösparti J. – Szabó J. – Bakacsi Zs. – Kuti L. 2006. Spatial stochastic modelling of inland inundation hazard. 9th AGILE Conference on Geographic Information Science. Visegrád, Hungary
- Pradhan B. – Lee S. 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental, Modelling & Software* 25: 747-759
- Rakonczi J. – Mucsi L. – Szatmári J. – Kovács F. – Csató Sz. 2001. A belvizes területek elhatárolásának módszertani lehetőségei. I. Földrajzi Konferencia, Szeged
- Rakonczi J. – Csató Sz. – Mucsi L. – Kovács F. – Szatmári J. 2003. Az 1999. és 2000. évi alföldi belvív-elöntések kiértékelésének gyakorlati tapasztalatai. *Vízügyi Közlemények* 1998-2001. évi árvízi külön füzetek Vol. 4: 317-336
- Retter, Gy. 2006. Fuzzy, Neurális Genetikus, Kaotikus Rendszerek. Budapest: Akadémiai Kiadó. 425 p
- Sárközy F. 1998. Mesterséges neurális hálózatok mint GIS függvények. *Geomatikai Közlemények* 1: 109-130
- Tarboton D. G. – Bras R. L. – Rodriguez-Iturbe I. 1991. On the Extraction of Channel Networks from Digital Elevation Data. *Hydrological Processes* 5: 81-100
- Tobak Z. – Kitka G. – Szatmári J. – van Leeuwen B. – Mucsi L. 2008. Kiszélesített, Kiszélesített (SFAP) CIR légifelvétel készítése, feldolgozása és alkalmazása környezeti vizsgálatokban. IV. Magyar Földrajzi Konferencia, Debrecen
- Zurada J. M. 1992. Introduction to Artificial Neural Systems. New York: West Publishing Company. 683 p