Assessment of predictive ability of Artificial Neural Networks using Holographic Mapping

András Tompos[a*], József L. Margitfalvi[a], Lajos Végvári[b], Ernő Tfirst[a]

[a] Institute of Surface Chemistry and Catalysis, Chemical Research Center, Hungarian Academy of Sciences, 1525 Budapest, POB 17, Hungary

[b] Meditor General Innovation Bureau, 2623 Kismaros, Zrinyi u. 16., Hungary

Abstract

In this study Artificial Neural Networks (ANNs) have been used to reveal quantitative relationship between catalytic composition and catalytic activity. This relationship has been predefined using a hypothetical experimental space described by a multidimensional polynomial. The predictive ability of ANNs has been investigated, i.e. an attempt was done to evaluate how ANNs can envisage a given hypothetical experimental space. Data sets for training, validation and testing of ANNs have been obtained from the hypothetical experimental space using two different ways of sampling. Data were selected (i) by means of our optimization algorithm called Holographic Research Strategy (HRS), and (ii) randomly. In order to model real experimentation, data were also generated with error. The relationship between the complexity of different network topologies and their predictive ability has been investigated. It has been shown that when data used for training have been perturbed with a given level of noise, less complex network architectures give acceptable accuracy. Additionally, estimated experimental spaces were visualized in a 2D layout by means of Holographic Mappings (HMs). Analysis of HMs revealed that ANNs trained by data sets obtained upon an optimization procedure provides better description of the experimental

[*] Corresponding author: Tel: +36-1-4381163: fax: +36-1-4381143
E-mail address: tompos@chemres.hu

space in the vicinity of the optimum than ANNs trained by randomly selected data sets. This fact indicates again the importance of the optimization in combinatorial catalyst library design.

Keywords: multidimensional space, high-throughput methods, knowledge extraction, information mining, Artificial Neural Networks (ANNs), visualization, library optimization

Introduction

In high-throughput experimentation huge amount of data are created. In order to accelerate the process of optimization and discovery of new materials these data have to be analyzed by different information mining tools. The analysis of patterns between input and output variables is usually called information mining. Eventually, on the bases of data accumulated it has to be established how the input variables have to be selected in order to get higher system performance. Non linear dependence of objective functions (output variables) on the input variables, which is most common in many engineering disciplines, leads to complicated hyper surfaces that have to be optimized. However, it has to be emphasized that besides quantitative variables there are qualitative ones as well, such as for example the type of support, way of catalyst preparation, way of pretreatment etc. These variables create a set of hyper surfaces. The set contains *n* hyper surfaces with *n* equals the multiplication of all modalities of all qualitative variables.

Information mining is essential for rapid discovery of novel materials including heterogeneous catalysts. Fast and reliable convergence to the optimum is based on reliable approximation of the objective function with a model properly fitted to previously accumulated data. In our previous work it was shown that combined application of Artificial Neural Networks (ANNs) and Holographic Research Strategy (HRS) is highly efficient in information mining and knowledge extraction [1].

Practically, ANNs can be considered as black boxes relating input data to output ones. The black box character is due to the complex structure of neural networks. These systems mimic the mechanism acting in biological neuron networks. In Multilayer Perceptrons (MLPs), which are the most widely used type of ANNs in high-throughput materials research, including heterogeneous catalysis [1-7], the input signals propagate from layer to layer containing different number of neurons (nodes) resulting in finally the output signals. ANNs

are excellent for non-linear regressions. Their pattern recognition and predictive ability make them especially beneficial in different optimization tasks, such as optimization of heterogeneous catalysts libraries [1-7]. Based on input variables, such as composition, preparation and testing conditions etc., the catalytic performance of a given material can be predicted. Fitting the parameters of ANNs, usually called training, requires previously accumulated data of the related catalytic reaction. The set of catalytic data used in ANN training forms the so-called training set. A validation set is employed also in order to avoid overtraining of ANNs.

Because of their black box character, ANNs are not easy to be conceptualized. They cannot be used alone for optimization. They do not propose new samples to be tested according to an optimization criterion. Therefore, when optimization is performed ANNs are usually combined with different optimization algorithms such as Genetic Algorithm (GA) [4-6]. In such combination the optimization algorithm generates virtual compositions that have to be tested by ANNs. In virtual catalytic experiments repeated application of virtual preparation and testing steps leads to virtual hits (promising catalysts). In this approach the performance of virtual hits has to be tested in real catalytic experiments as well.

In the present study HRS [1] has been used as an optimization algorithm. It has been realized that besides the excellent optimization feature of HRS it has singular visualization ability, too. Holographic Mapping (HM) can represent multi-dimensional surfaces, i.e. both multi-dimensional functions and multi-dimensional experimental spaces, in a 2D layout [1,8]. The combination of HM with ANNs provides not only a virtual optimization system, but a visualization tool as well [1]. In this way relationships in multi-dimensional spaces can be revealed; complex hyper surfaces can be analyzed visually. Eventually, HRS transforms sophisticated mathematical structure of ANNs into a human-comprehensible representation. Practically, HM enlightens the black box character of the ANNs, consequently the

combination of HM with ANNs appears to be an excellent method for knowledge extraction [1].

It has to be mentioned that beyond ANNs other machine learning techniques, such as for example Support Vector Machines (SVM) [9,10], are also well known. They are supposed to be superior to ANNs due to significantly less tendency to be overfitted. However, according to our opinion their generalization and predictive ability is not better than that of ANNs. The comparison of SVM to ANNs trained according to our approach and methods (vide infra) can be a subject of another study.

In most of the case studies using combinatorial approaches in heterogeneous catalysis the catalytic performance (conversion of substrate, product selectivity or yield, overall activity of the catalysts) has been investigated as the function of catalysts composition [1-5]. Nevertheless, there were attempts to include other parameters, such as reaction temperature, type of precursor compounds and mode of preparation [6,7,11]. However, according to our opinion nominal variables (types of precursors used, preparation methods and pretreatment conditions applied, etc.) are advisable to be kept constant in the process of design of a given catalyst library. Any alteration leads to a new experimental space (a new hyper surface) that has to be optimized separately. There can be significant discontinuity at the interfaces between the different experimental spaces, which cannot reasonably be treated with any optimization algorithms. In this case study only the composition of the catalysts is altered.

The assessment of the predictive ability of neural networks is a crucial step in making decision how they can be used in further optimization. The relationships provided by properly trained ANNs can be used for extrapolation, i.e. to predict catalytic performance over the domain of independent variables tested so far.

In the process of evaluation of neural networks, predicted performance and real experimental data have to be compared. It is usually based on the analysis of a testing set that

contains experimental data that were not included either in the training set or the validation set. The comparison of measured values to the predicted ones can be done calculating the Mean Square Error (MSE) or representing the predicted-measured correlation in a line graph [1,2]. In case of small testing set cross validation or boostrap can also be used [9]. Nevertheless, it has to be emphasized that the test set always contains arbitrarily selected catalysts. For example, upon applying an effective optimization algorithm the bad performing catalysts are discarded and in this way bad experimental regions remain under-sampled. Consequently, the training and test sets contain catalyst compositions predominantly from the vicinity of the optimum. Therefore it is highly probable that this experimental region is described more accurately than experimental regions of poor catalysts. A good correlation between measured and predicted data is generally due to the fact that the data are taken from the same class of quality. However, this good correlation is misleading *since the under-sampled part of the experimental space was not investigated and therefore cannot be estimated adequately.*

In the present study in order to assess the predictive ability of trained neural networks, a hypothetical experimental space is applied involving multiple independent variables in a multidimensional polynomial adopted from our previous paper [8]. Consequently, in the benchmark the composition-activity relationship is known. The comparison of "measured" (calculated) and predicted values is not restricted to an arbitrarily selected set of catalysts as the whole experimental space is available. The experimental space containing the calculated data will be called as "original" experimental space while that one containing the values predicted by neural networks will be called as "estimated" experimental space.

Holographic Mapping will be applied for the comparison of original and estimated experimental spaces. Benchmark analysis, i.e. the application of a multidimensional function that describes explicit relationship between the composition and the catalytic performance

makes possible to perform unambiguous evaluation of ANNs and thus the conclusions can be generalized and exploited later in real experimentation.

It has to be mentioned, that the strength of ANNs has already been demonstrated in many fields. Therefore, in this study the focus will be laid on applicability of Holographic Mapping as a new tool to evaluate the prediction ability of neural networks using a predefined benchmark. It will be shown that the application of MSE does not allow evaluating directly the predictive ability of ANNs. Thus, local visual analysis is required, i.e. the square errors have to be visualized in Holographic Maps. In this paper the effects of three crucial parameters, such as (i) sampling method, (ii) ANNs architecture, and (iii) experimental error, on the predictive ability of ANNs are investigated.

First of all, two different sampling methods are tested. Catalysts will be selected from the original experimental space (i) upon using an optimization process (HRS), and (ii) randomly. In the previous case the density of data around the optimum is significantly larger than in the latter one, which is expected to affect the predictive ability of ANNs trained with these two different data sets.

The effect of the architecture of the ANNs will also be investigated on the predictive ability of neural networks. In case of MLPs, the effect of the number of both hidden layers and hidden neurons in different layers will be studied.

Thirdly, it is known that real experimental data always contain error. In the training phase neural networks are capable to learn the noisy data sets as well, which is supposed to affect the predictive ability of ANNs. In order to model experimental error, data in the original experimental space are modified randomly in a ±5 % interval. Both noisy and noise-free data sets are used in individual training sets.

Methods

Hypothetical objective function

In this study a hypothetical experimental space is applied modeling multi-component supported metal catalysts adopted from our previous paper [8]. The oxidation of CO over hypothetical Pt and Pd based multi-component catalysts has been chosen as a virtual reaction. The objective function according to formula (1) has been used to describe the yield of CO_2 (Y_{CO2}) as a function of metal content (given in wt %).

$$\mathbf{Y_{CO_2} = 4.5 \cdot Pt + 0.5 \cdot Pd + 0.5 \cdot Ru + 0.5 \cdot Pt \cdot Pd \cdot (Pt+1) \cdot (Pd+0.1) + Pt \cdot Sn \cdot (Pt-0.8) \cdot (Sn+0.8) + Pt \cdot Pd \cdot Sn \cdot (Pt+Pd) \cdot (Sn+1) + Ru \cdot Sn \cdot (Sn+1) \cdot (Ru-0.1) - 0.1 \cdot (Ge+In) \cdot (Pt+Pd+Ru)} \quad \mathbf{(1)}$$

Holographic Research Strategy

HRS is an optimization algorithm, which additionally has unique visualization abilities, too [1,8]. Holographic Mapping provides integrated overview of multidimensional experimental spaces or multidimensional functions in a two-dimensional form. Holographic Mapping is described in detail in our previous study [8]. Variables are arranged along the perpendicular X and Y-axes. Lines with different length substitute the discrete levels of different variables (see e.g. Fig. (**1**)). It has to be emphasized that a given line represents one level of a variable and the length of a line is proportional to the number of data points displayed along the line. The visualization is based on wavelike arrangement of levels of independent variables along the X and Y-axes. The level of each component increases gradually till it reaches its maximum then it decreases gradually again (see Fig. (**1**)). Accordingly, moving along any axis from one experimental point to the next one only one level of one variable is changed. This arrangement results in a 2D matrix where the adjacent points are neighbors in the original multidimensional space as well.

In the present paper both the original hypothetical experimental space and different experimental spaces estimated with different neural networks (see later) have been visualized

using the Holographic Mapping. Additionally, the values of square errors representing the difference between corresponding data points in the 2D matrix of the original and estimated experimental spaces have also been visualized. These maps will be called as square error maps.

The eight catalyst components and their concentration levels are summarized in Table 1. The total combination of all concentration levels leads to an experimental space containing 81,920 simulated CO_2 yield data. In HMs the first four variables are arranged along the X-axis while the second four variables along the Y-axis (see Fig. (**1**)).

In order to fit the parameters of neural networks three sets of catalysts (training, validation and testing sets) have to be selected from the hypothetical experimental space. It has been done in two different ways: catalysts are selected (i) according to HRS optimization and (ii) randomly. In the former case optimization steps described in our previous studies have been applied [1,8].

In HRS optimization the initial experimental points (first catalyst generation) have been fixed as small clusters (i) in the center, (ii) along the symmetry axes and (iii) in the corners of the HMs of experimental spaces resulting in 48 different catalyst compositions [1,8]. The forthcoming generations have been created by a rectangular-shaped experimental region 4x4 in size around the best three hits of the preceding generations. Prior to the formation of the experimental regions a variable position change has to be done according to our previous routines. This step is considered as the main driving force in HRS. As a result of variable position changes the arrangement of compositions in the experimental hologram is altered, eventually after variable position changes each hit has a new neighborhood [1,8].

Artificial neural networks

The formation of ANNs requires previously measured experimental data. In the present study "yields" are calculated using formula 1. Four different data sets, each containing about 200 data, have been created according to the combination of two different sampling methods (HRS optimization and random) and two different experimental errors (0 % and 5 %).

For proper formation of ANNs and for the assessment of their predictive ability the yields have been divided into three well-distinguishable sets: (i) training, (ii) validation and (iii) testing. In this study the ratio of training, validation and testing sets was adjusted to approximately 70:15:15, respectively.

The process of training and validation has been described in detail elsewhere [3]. The networks are trained with resilient back-propagation algorithm [3]. Training is stopped if the validation error increases for more than two consecutive epochs; an epoch is defined as a pass through the entire set of training and validation patterns. This protocol is invoked to prevent over-teaching during the training phase [3].

Nineteen different kinds of network architecture (see Table 2) proposed by Cundari et al. [3] have been investigated to achieve acceptable model accuracy. Complexity of different networks used in this study is reflected in the number of adjustable connection weights, which is dependent on the number of hidden layers and number of hidden neurons in different layers. The first network is the simplest; it consists of only one hidden layer and only 5 hidden neurons in this layer, which results in a topology only with 51 connection weights. In Table 2 the applied neural networks are ranked according to the increasing number of their connection weights.

Usually, the selection of an appropriate architecture is problem dependent. The approach proposed by Cundari makes possible an automatic selection of topologies. Actually,

a so-called Optimal Linear Combination [12] of different networks is used instead of application of a single topology. The combination weights for different networks are computed using the ordinary least mean square error algorithm and the statistically insignificant (at a 0.05 level of significance) architectures have been discarded.

Before linear combination, each kind of architecture has been trained 1000 times (each training has been initialized with different, random node-to-node weights) [3], which results in 1000 networks from each topology. According to the 19 different topologies applied, the whole set consists of 19000 networks. In order to investigate the predictive ability of different topologies, not the whole set was used but networks with different complexity have been selected from Table 2. For example, in case of the first two topologies, prior to linear combination, the initial set consists of 2000 networks only. This number was further reduced to the best 100 networks. They are considered as the best, since they have the smallest mean square errors (MSEs) calculated for both training and validation sets. Finally, the best 100 networks have been involved into the Optimal Linear Combination [12] mentioned above, during which a so-called OLC-network has been created. The linear combination can also be considered as a final reducing step, upon which the statistically insignificant networks are removed.

Similarly to the strategy described above three additional OLC-networks have been created using the first four, eight and nineteen network topologies, respectively. Eventually, the resulting four OLC-networks have been applied in this study in combination with HRS in order to get a map of estimated experimental spaces. OLC-networks are designated as ANN2, ANN4, ANN8 and ANN19 indicating the numbers of network architectures involved into the linear combinations.

In the designation of networks the abbreviation “opt” means that the data set has been created upon using the optimization algorithm (HRS), while “rnd” refers to the random

selection of data. The last number in the designation of the OLC-network indicates the value of generated experimental error. For example, a network designated as ANN8_opt_5 refers to an OLC-network that was created using the first eight types of architecture trained with data set obtained after optimization with HRS and generated with 5 % experimental error.

Results and Discussion

Holographic Mapping of the original experimental space is given in Fig. (**1**). Data have been generated according to equation (1). As emerges from Fig. (**1**) the preferential role of Pt can be revealed. Dark points are concentrated at high Pt content. Moreover, a cross effect between tin and platinum can also be recognized as highly active compositions can only be observed when both Sn and Pt are present. Detailed analysis of the original experimental space has been given in our earlier study [8].

This experimental space has been estimated by means of ANNs. This was done by using four different data sets according to two different sampling methods (HRS optimization and random selection) and two different extents of generated experimental errors (0 % and 5 %). Based on the above data sets neural networks have been trained and OLC-networks have been created. In Tables 3-6 different OLC-networks obtained upon using the first two, four, eight and nineteen network topologies, respectively are characterized. It is remarkable, that in spite of that the best 100 trained networks have been involved into the linear combinations only a few proved to have combination weights significantly different from zero. For example, in Table 3 the OLC-network ANN2_rnd_0 consist of only 3 component networks having 0.15, 0.56 and 0.30 combination weights, respectively. The rest 97 networks have been dropped out since their combination weights have been insignificant (at a 0.05 level of significance).

It has to be mentioned that Tables 3-6 contain two types of MSE values. MSE_{full} has been calculated considering all data in the full experimental space (data used for training and validation were excluded from these calculations). MSE_{test}, has been determined according to the conventional approach mentioned in the Introduction, i.e. using the test set only. In our particular case it was restricted only to 27-33 data. As emerges from Table 3-6, the two types of MSE are not fully correlated and in our further analysis we have attributed higher relevance to MSE_{full}, which might reflect the predictive ability of trained neural networks more reliable than the MSE_{test}.

Effect of network architecture on the predictive ability of OLC-networks

As merges from Tables 3-6 the complexity of neural networks involved in the linear combinations increase in the following order: of ANN2<ANN4<ANN8<ANN19. For example, in case of ANN19_rnd_0 a three-hidden-layered component network can also be found (see Table 6). It is worth emphasizing again that the training procedure resulted in 1000 networks from each kind of architecture, i.e. in case of ANN19 altogether 19000 networks are created. Hence, originally the less complex networks with low number of hidden layers and hidden neurons were available as well. According to the first reduction step from these nineteen thousand networks the best 100 have been selected for linear combinations. The linear combination was applied as a second reduction step, which eventually resulted in the 6 component networks for ANN19_rnd_0 as shown in Table 6. All these steps are responsible for the decreased number of networks with low number of hidden layers and connection weights, although originally the less complex networks were present as well. Apparently, the system described by formula (1) requires the formation of complex neural architectures.

If the system really requires complex network architecture, the MSEs obtained with respect to the “original” experimental space are expected to decrease in the following order:

ANN2>ANN4>ANN8>ANN19. However, this assumption is only true when error free data set was used for the training (see Table 3-6 and Fig. (**2**)). In comparison of ANN2_opt_0 to ANN19_opt_0 the MSE decreases from 8.20 to 4.78, respectively. However, when the data set is generated with 5 % error the MSE decreased only in the first three OLC-networks (ANN2, ANN4 and ANN8) and it increased again when ANN19 was applied (see Fig. (**2A**)). The results show that the nonlinear character of the system is masked by the experimental noise and therefore a less complex network can give as good prediction as a more complex one.

It can be mentioned that similar results were obtained when other hypothetical functions with multiple variables have been applied. Tens of different multidimensional functions have been tested, but detailed discussion of these results exceeds the frame of the present study. Nevertheless, the conclusion is always the same: noisy data sets, which are very frequently generated in real experimentation, require training of less complex networks.

It has to be emphasized that the MSE_{full} between original and estimated data have been calculated considering all data in the benchmark, of course, with the exception of data used for training and validation. Intrinsically, this value shows only an average picture about the predictive ability of an OLC-network. Square error can significantly be changed in different regions of the experimental space (vide infra).

Hereinafter the effect of experimental error (0 and 5 %) and the sampling method (via optimization and random selection) on the predictive ability will be analyzed using ANN8-networks.

Effect of sampling method and experimental error on the predictive ability of OLC-networks

The conventionally used line graphs, which are plotted as predicted values vs. measured ones using data of test sets, are shown in Fig. (**3**)-(**4**) obtained in case of the four ANN8-networks. In Fig. (**3**) the applied OLC-networks have been created by data set obtained via optimization by means of HRS. The training set was generated both without and with 5 % error in case of ANN8_opt_0 and ANN8_opt_5, predictions of which are shown in Fig. (**3A**) and (**3B**), respectively. In case of random selection of data sets the corresponding line graphs are shown in Fig. (**4**). The MSE_{test} values calculated with respect to the test sets are listed in Table 5.

It has to be mentioned that prior to training, data with different quality are distributed evenly in training, validation and test set so that in each set (training, validation and test) all sorts of catalysts from the poorest to the best available after different data selection methods, are used with the same rate. Therefore, the test sets applied in Fig. (**3**)-(**4**) reflect the data distribution in the other two sets (training and validation) as well. Fig. (**4**) clearly shows that the test sets do not contain higher yield than 30 %. It means that during the random selection of data the best CO_2 yield obtained was not higher than 30 %. Neither the training set nor the validation set can contain higher yield data. In contrast to this in Fig. (**3**) a broader distribution of data can be seen, i.e. during the HRS optimization, data with higher CO_2 yield were also found and thus were involved into the test set as well. As emerges from Fig. (**1**) the original experimental space mostly contain data below 24 % CO_2 yield. Therefore, it is not a surprise that by random selection high yields could not be found.

At the first sight, preferential role of random selection of data to selection by optimization seems to be evident. In Fig. (**4**), where ANN8_rnd networks were applied, excellent correlation between the predicted and measured values is achieved. The corresponding MSE_{test} values calculated with respect to the test set (see Table 5) were 0.10 and 2.94 at ANN8_rnd_0 and ANN8_rnd_5, respectively. In contrast to this, Fig. (**3**) indicates significantly higher deviations upon using ANN8_opt networks. The MSEs are 5.71 and

16.00 at ANN8_opt_0 and ANN8_opt_5, respectively. Additionally, in case of ANN8_rnd networks MSEs obtained with respect to the full experimental space (see Fig. (**2B**)) are significantly lower than those obtained in case of ANN8_opt ones (see Fig. (**2A**)). However, these observations can be misleading in the assessment of the predictive ability of OLC-networks. In case of line graphs (see Fig. (**3**) and (**4**)) the comparison between the predicted and measured values are restricted to a very limited number of data, which are additionally far away from the optimum, while in case of MSE_{full} values (see Fig. (**2**)) the alteration of the square error throughout the experimental space cannot be taken into consideration. *Therefore, holographic maps and the so-called square error maps have to be applied in order to get a clear picture about the predictive ability of different OLC-networks (These results are shown in Figs. (**5**)-(**8**)).*

Experimental spaces estimated by ANN8_opt_0 and ANN8_rnd_0 (networks trained by data without noise) are given in Fig. (**5A**) and Fig. (**7A**), respectively. The corresponding square error maps are shown in Fig. (**5B**) and Fig. (**7B**), respectively. In comparison to the original experimental space (see Fig. (**1**)) it can clearly be seen that the structure of data was estimated in a somewhat better way when neural networks were trained with data selected randomly (ANN8_rnd_0). However, as emerges from square error maps (Fig. (**5B**) and Fig. (**7B**)) the ANN8_opt_0 provides a much better approximation of the original experimental space in the vicinity of the optimum than the ANN8_rnd_0. The upper right side of Fig. (**5B**) is homogeneously light gray indicating small square error when ANN8_opt_0 was used, whereas the same region in Fig. (**7B**) reveals dark spots referring to bad estimation of the experimental space when ANN8_rnd_0 was used. In contrast to this, ANN8_rnd_0 proved to be much efficient in prediction at low CO_2 yield data than ANN8_opt_0. As the original experimental space mostly contains data below 24 % CO_2 yield, the MSE_{full} values calculated in case ANN8_rnd_0 can be significantly lower than in case ANN8_opt_0. Nevertheless,

from point of view of an experimenter the better prediction around the optimum possesses much higher significance than in any other regions of the experimental space. It is the main message from the above results.

Beyond the above-mentioned correlation between noisy data sets and selected architecture, the use of data with experimental error results in another obvious consequence. As emerges from Table 3-6 and Fig. (**2**) the MSEs are significantly smaller when data without noise are used for training. The accuracy of the data set applied in training determines fundamentally the accuracy of predictions. Generally, a model, parameters of which is fitted by using noisy data set, cannot predict data more accurately than the accuracy of applied noisy data set itself. In this case the mean square error is a result of superposition of both the experimental and the prediction errors. Nevertheless, as emerges from Fig. (**6**) and Fig. (**8**) the estimation of experimental spaces are acceptable even if noisy data sets were used for the training of neural networks. It can be added that in high-throughput experimentation, similarly to other experimental works, the appearance of error is unavoidable; therefore a model that can tolerate noisy data is highly desirable. According to our result ANNs provide appropriate model even if the training set is noisy.

Summary

In this study a multidimensional function that describes relationship between the composition and the catalytic performance has been applied in order to generate a hypothetical experimental space. Data selected from this space were used in training of ANNs. The predictive ability of ANNs has been investigated, i.e. we have tried to answer the question: how ANNs can estimate an exactly defined hypothetical experimental space. The application of MSEs in the evaluation of ANNs can be misleading because it is either restricted only to a selected test set or provides only an average value throughout the whole experimental space.

Contrary to this, the local visual analysis of square errors in holographic maps can reveal the predictive ability of different neural networks.

It can be established that the presence of experimental noise in the training set influences significantly the complexity of requested neural architecture. ANN8-networks have consisted of mostly one-hidden-layered component networks only and the corresponding calculated MSEs have been comparable to those obtained when ANN19-networks with two-hidden-layered component networks were applied (see Table 5 and 6). It could be concluded that the nonlinear character of the system is masked by the experimental noise and therefore a less complex network can give as good prediction as a more complex one. Additionally, analysis of the holographic maps leads to the conclusion that ANNs provide acceptable prediction even if they were trained with noisy data sets (see Fig. (**6**)). Therefore, application of ANNs in real experimentation is reasonable and useful.

It was also shown that neural networks trained with data set obtained after holographic optimization provide better model accuracy around the optimum, than those trained with randomly selected samples. This fact has again a great importance in real experimentation, especially when the goal is to find optimal compositions. In this way, alternating application of real optimization and neural networks can lead to accelerated optimum search.

Acknowledgment

Partial financial support given to A.T. (OTKA grant N^{o} F 049742) is greatly acknowledged.

References

[1] Tompos, A.; Margitfalvi., J.L.; Tfirst., E.; Végvári., L.; Jaloull., M.A.; Khalfalla., H.A.; Elgarni., M.M. *Appl. Catal. A: Gen.,* **2005**, *285*, 65.

[2] Holeňa, M.; Baerns, M. *Catal. Today,* **2003**, *81* 485.

[3] Cundari, T.R.; Deng, J.; Zhao, Y. *Ind. Eng. Chem. Res.,* **2001**, *40*, 5475.

[4] Huang, K.; Zhan, X.L.; Chen, F.Q.; Lü, D.W. *Chem. Eng. Sci,.* **2003**, *58*, 81.

[5] Corma, A.; Serra, J.M.; Argente, E.; Botti, V.; Valero, S. *Chem. Phys. Chem.,* **2002**, *3* 939.

[6] Baumes, L.A; Farrusseng, D.; Lengliz, M.; Mirodatos, C. *QSAR & Comb. Sci.* **2004**, *23*, 767.

[7] Farrusseng, D.; Klanner, C.; Baumes, L.A.; Lengliz, M.; Mirodatos, C.; Schüth, F. *QSAR & Comb. Sci.* **2005**, *24*, 78.

[8] Végvári, L.; Tompos, A.; Gőbölös, S.; Margitfalvi, J. *Catal. Today*, **2003**, *81*, 517.

[9] Baumes, L.A.; Serra, J.M.; Serna, P.; Corma, A. *J. Comb. Chem.,* **2006**, *8*, 583.

[10] Baumes, L.A; Moliner, M.; Corma, A. C. *QSAR & Comb. Sci.,* **Published Online: 30 Aug 2006***,* DOI: 10.1002/qsar.200620064.

[11] Serra, J. M.; Corma, A.; Farrusseng, D.; Baumes, L.; Mirodatos, C.; Flego, C.; Perego, C., *Catal. Today*, **2003**, *81*, 425.

[12] Hashem, S. *Neural Networks* **1997**, *10*, 599.

Figure captions

Fig. (1). Holographic mapping of "measured" (calculated) CO_2 yield data in the "original" experimental space. Color-codes of yield data are indicated in the figure. The sequence of variables with increasing frequency of compositional waves along the X- and Y-axes: Pt, Pb, Sn, In and Ga, Ge, Ru, Pd, respectively.

Fig. (2). Mean square errors obtained upon using different OLC-networks trained by data set obtained either in an optimization process by means of HRS or randomly (A and B, respectively). Effect of OLC-networks is compared using data sets ♦ without error and O with 5 % generated error. The MSE between the original and estimated data has been calculated considering all data in the benchmark with the exception of data of training and validation set.

Fig. (3). Prediction ability of OLC-networks trained with the data sets obtained in optimization process by means of HRS (ANN8_opt) generated without error and with 5 % error (A and B).

Fig. (4). Prediction ability of OLC-networks trained with the data sets selected randomly (ANN8_rnd) generated without error and with 5 % error (A and B).

Fig. (5). Holographic mapping of predicted CO_2 yield and square errors data (A and B) in the estimated experimental space by means of OLC-Network ***ANN8_opt_0***. Data are presented in grayscale indicated in the figure. For designation of components see Fig. (**1**).

Fig. (6). Holographic mapping of predicted CO_2 yield and square errors data (A and B) in the estimated experimental space by means of OLC-Network ***ANN8_opt_5***. Data are presented in grayscale indicated in the figure. For designation of components see Fig. (**1**).

Fig. (7). Holographic mapping of predicted CO_2 yield and square errors data (A and B) in the estimated experimental space by means of OLC-Network ***ANN8_rnd_0***. Data are presented in grayscale indicated in the figure. For designation of components see Fig. (**1**).

Fig. (8). Holographic mapping of predicted CO_2 yield and square errors data (A and B) in the estimated experimental space by means of OLC-Network ***ANN8_rnd_5***. Data are presented in grayscale indicated in the figure. For designation of components see Fig. (**1**).

Table 1. Concentration levels of the components of catalysts in the hypothetical experimental space

Levels	Components, in wt %							
	Pt	Pb	Sn	In	Ga	Ge	Ru	Pd
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.6	0.3	0.6	0.4	0.3	0.4	0.2	0.5
3	1.2	0.6	1.1	0.8	0.6	0.8	0.4	1.0
4	1.7	0.8	1.5	1.1	0.8	1.1	0.5	1.4
5	2.1	-	-	-	-	-	-	-

Table 2. Architecture of neural networks applied in this study (adopted from ref [3])

N^{o}	Number of hidden neurons in the hidden layers			Connection weights
	1st hidden layer	2nd hidden layer	3rd hidden layer	
1	5			51
2	5	3	2	74
3	5	5		81
4	10			101
5	15			151
6	20	1		201
7	25	1		251
8	10	15	1	271
9	14	10	1	287
10	15	10	1	306
11	30	1		301
12	35	1		351
13	15	15	1	391
14	20	10	1	401
15	40	1		401
16	45	1		451
17	20	10	10	500
18	50	1		501
19	20	20	1	623

Table 3. OLC-networks obtained in the training of the first two types of neural architecture

Designation of OLC-networks	Number of neurons in the hidden layers of component networks 1st	2nd	3rd	Combination weights[1]	$MSEs_{full}$[2]	$MSEs_{test}$[3]
ANN2_opt_0	5			0.54	8.20	11.32
	5			0.47		
ANN2_opt_5	5			0.58	26.75	15.52
	5			0.42		
ANN2_rnd_0	5			0.15	4.27	2.60
	5			0.56		
	5	3	2	0.30		
ANN2_rnd_5	5	3	2	0.36	4.46	1.08
	5			0.39		
	5			0.26		

[1] Coefficients in the linear combination of component networks [2] The mean square error between the original and estimated data considering all data in the full experimental space with exception the data of training and validation set. [3] The mean square error between the original and estimated data considering data in the test set.

Table 4. OLC-networks obtained in the training of the first four types of neural architecture

Designation of OLC-networks	Number of neurons in the hidden layers of component networks 1st	2nd	3rd	Combination weights[1]	$MSEs_{full}$[2]	$MSEs_{test}$[3]
ANN4_opt_0	10			0.31	8.98	14.61
	10			0.44		
	10			0.25		
ANN4_opt_5	10			1.00	25.08	31.65
ANN4_rnd_0	10			0.19	1.48	0.39
	10			0.25		
	5			0.18		
	10			0.39		
ANN4_rnd_5	10			0.44	3.80	2.16
	10			0.55		

[1] Coefficients in the linear combination of component networks [2] The mean square error between the original and estimated data considering all data in the full experimental space with exception the data of training and validation set. [3] The mean square error between the original and estimated data considering data in the test set.

Table 5. OLC-networks obtained in the training of the first eight types of neural architecture

Designation of OLC-networks	Number of neurons in the hidden layers of component networks 1st	2nd	3rd	Combination weights[1]	$MSEs_{full}$[2]	$MSEs_{test}$[3]
ANN8_opt_0	10			0.26	7.70	5.71
	25			0.22		
	20			0.20		
	25			0.32		
ANN8_opt_5	25			1.00	8.16	16.00
ANN8_rnd_0	10			0.12	1.04	0.10
	25			0.22		
	15			0.18		
	10	15		0.16		
	25			0.14		
	25			0.19		
ANN8_rnd_5	15			0.34	4.24	2.94
	10	15		0.46		
	5			0.20		

[1] Coefficients in the linear combination of component networks [2] The mean square error between the original and estimated data considering all data in the full experimental space with exception the data of training and validation set. [3] The mean square error between the original and estimated data considering data in the test set.

Table 6. OLC-networks obtained in the training all the nineteen types of neural architecture

Designation of OLC-networks	Number of neurons in the hidden layers of component networks 1st	2nd	3rd	Combination weights[1]	$MSEs_{full}$[2]	$MSEs_{test}$[3]
ANN19_opt_0	45			0.23	4.78	20.43
	50			0.27		
	15	15		0.25		
	45			0.26		
ANN19_opt_5	14	10		1.00	25.24	20.00
ANN19_rnd_0	14	10		0.24	1.59	1.46
	14	10		0.18		
	20	20		0.20		
	25			0.16		
	20	10		0.13		
	20	10	10	0.10		
ANN19_rnd_5	15	15		0.59	4.20,	3.06
	15	15		0.41		

[1] Coefficients in the linear combination of component networks.[2] The mean square error between the original and estimated data considering all data in the full experimental space with exception the data of training and validation set.[3] The mean square error between the original and estimated data considering data in the test set.

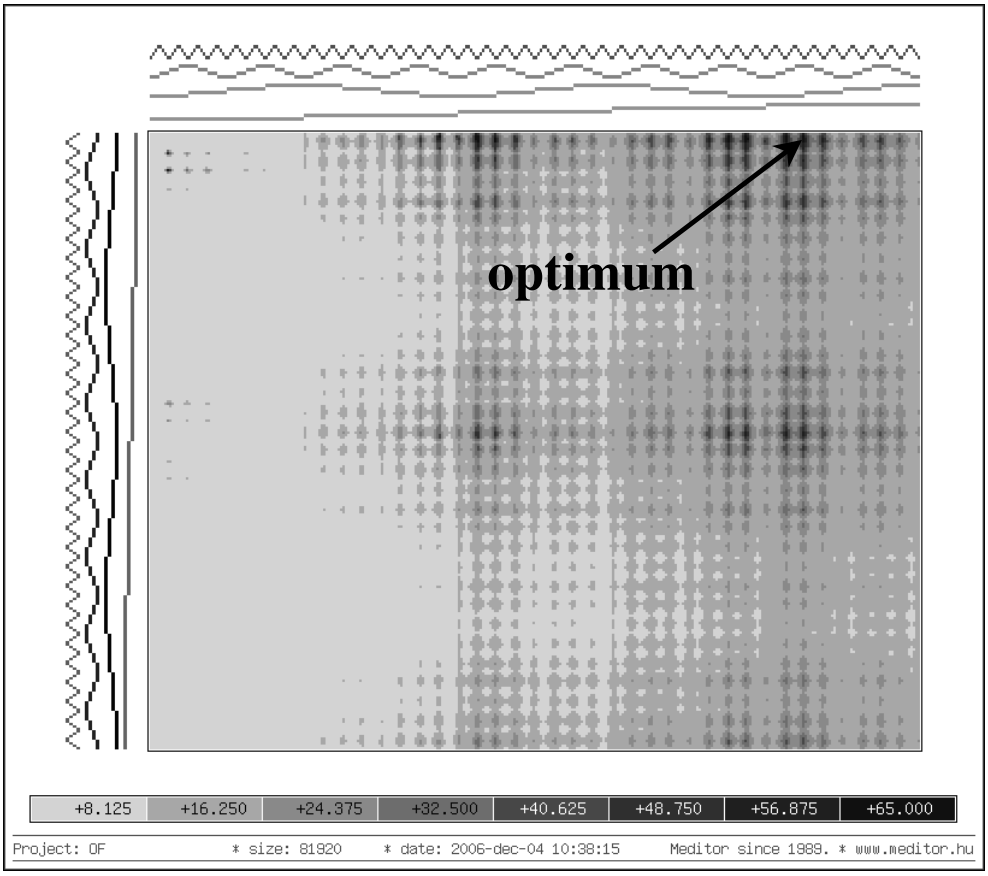
optimum
+8.125
+16.250
+24.375
+32.500
+40.625
+48.750
+56.875
+65.000
Project: OF
* size: 81920
* date: 2006-dec-04 10:38:15
Meditor since 1989. * www.meditor.hu

Fig. (1).

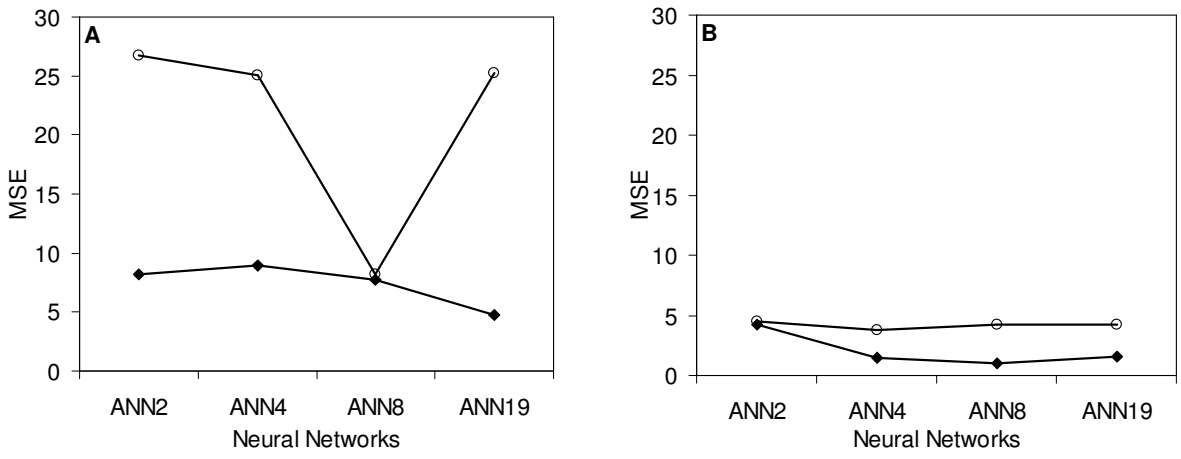

Fig. (2).

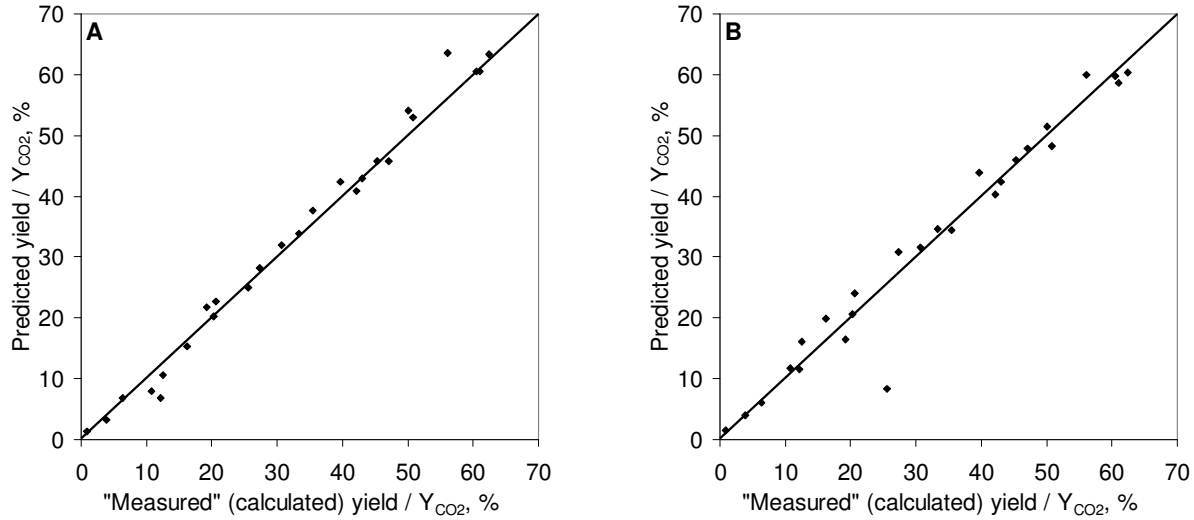

Fig. (3).

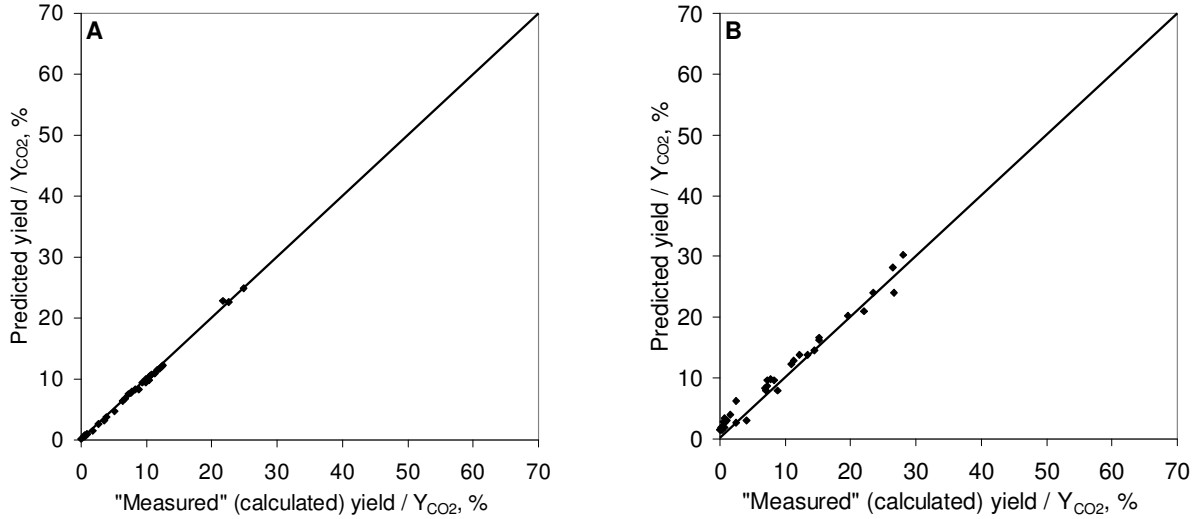

Fig. (4).

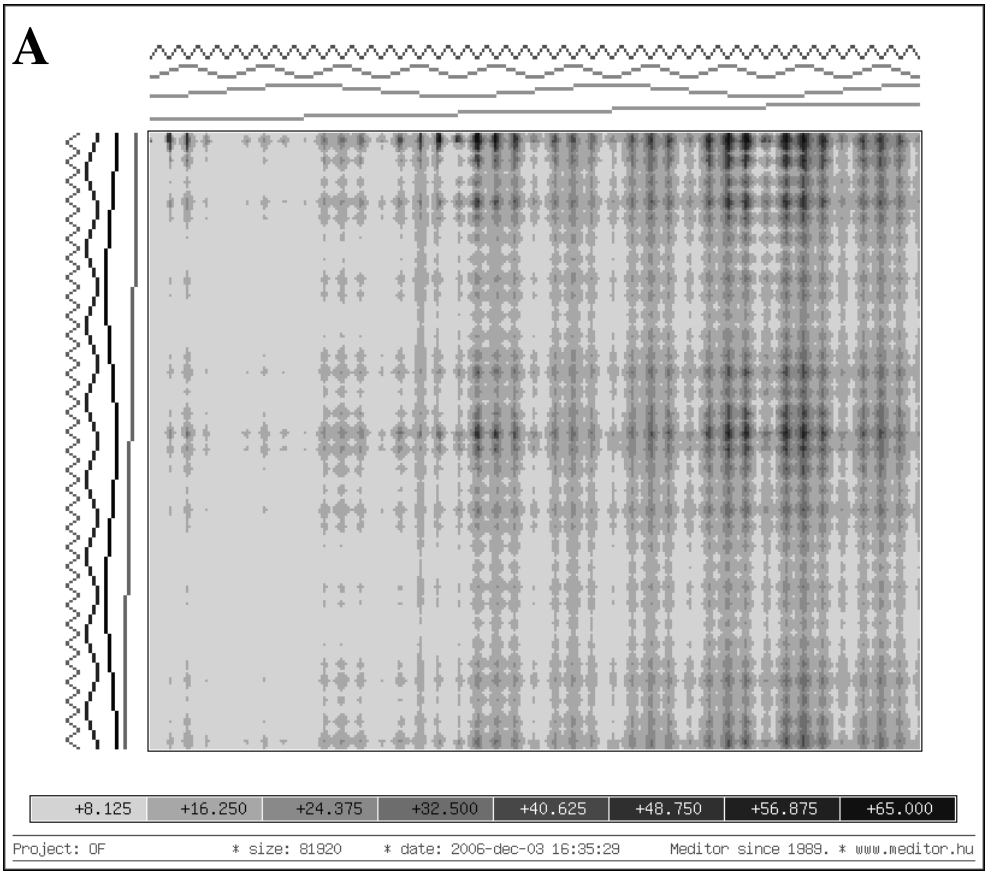
A
+8.125
+16.250
+24.375
+32.500
+40.625
+48.750
+56.875
+65.000
Project: OF
* size: 81920
* date: 2006-dec-03 16:35:29
Meditor since 1989. * www.meditor.hu

Fig. (5A).

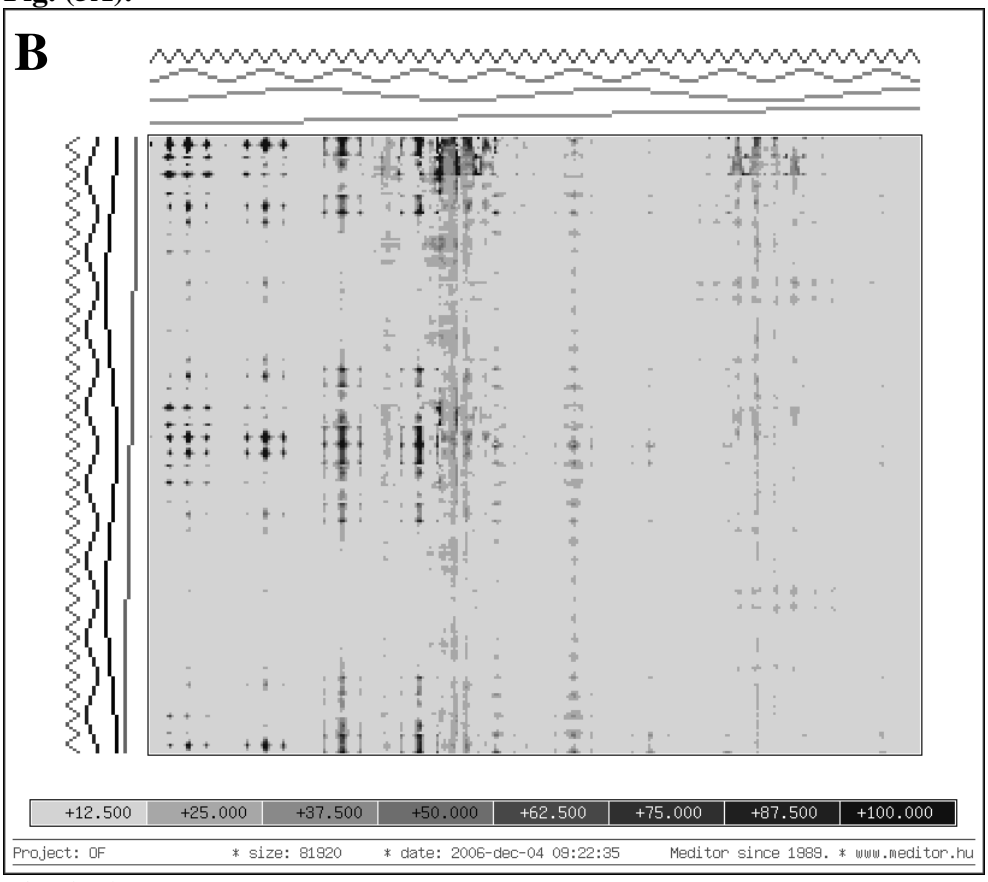
B
+12.500
+25.000
+37.500
+50.000
+62.500
+75.000
+87.500
+100.000
Project: OF
* size: 81920
* date: 2006-dec-04 09:22:35
Meditor since 1989. * www.meditor.hu

Fig. (5B).

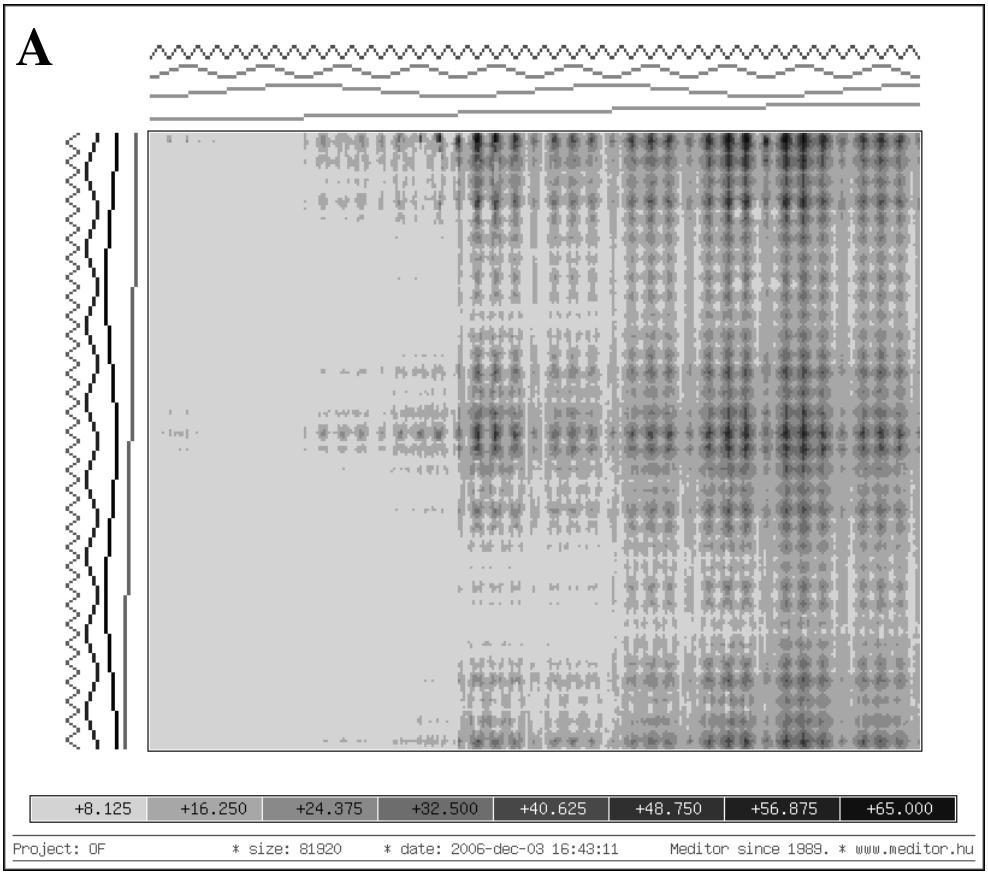
A
+8.125
+16.250
+24.375
+32.500
+40.625
+48.750
+56.875
+65.000
Project: OF
* size: 81920
* date: 2006-dec-03 16:43:11
Meditor since 1989. * www.meditor.hu

Fig. (6A).

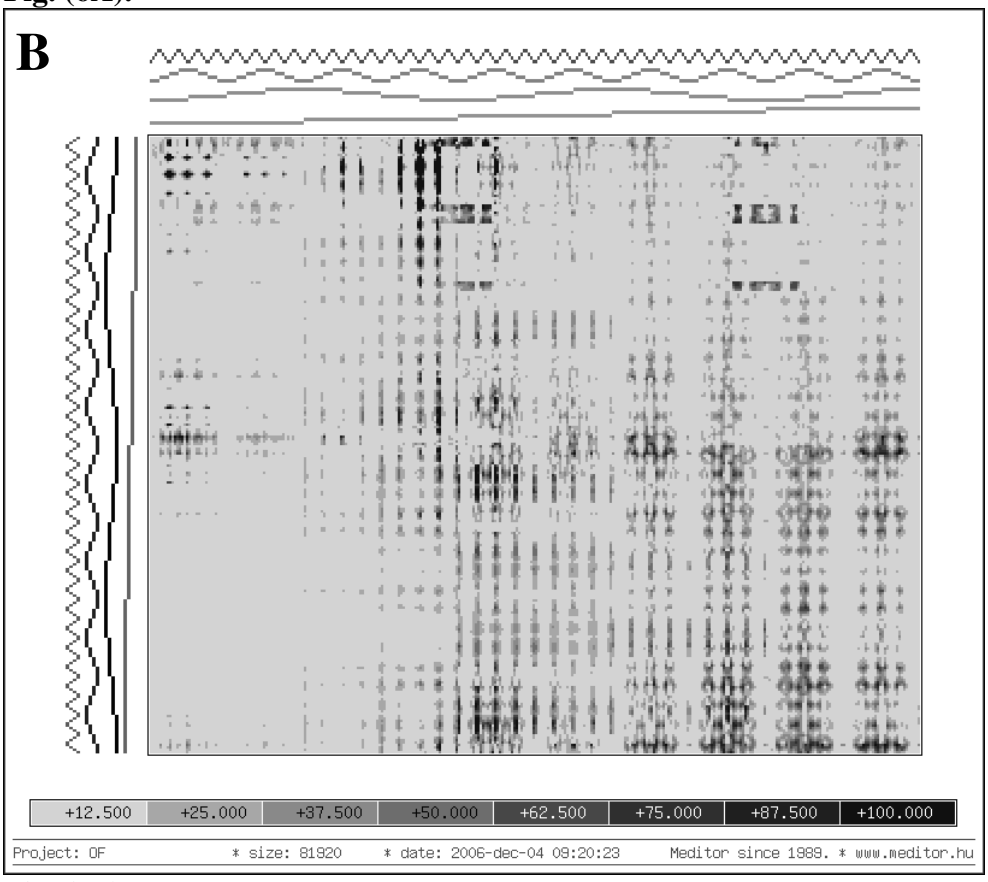
B
+12.500
+25.000
+37.500
+50.000
+62.500
+75.000
+87.500
+100.000
Project: OF
* size: 81920
* date: 2006-dec-04 09:20:23
Meditor since 1989. * www.meditor.hu

Fig. (6B).

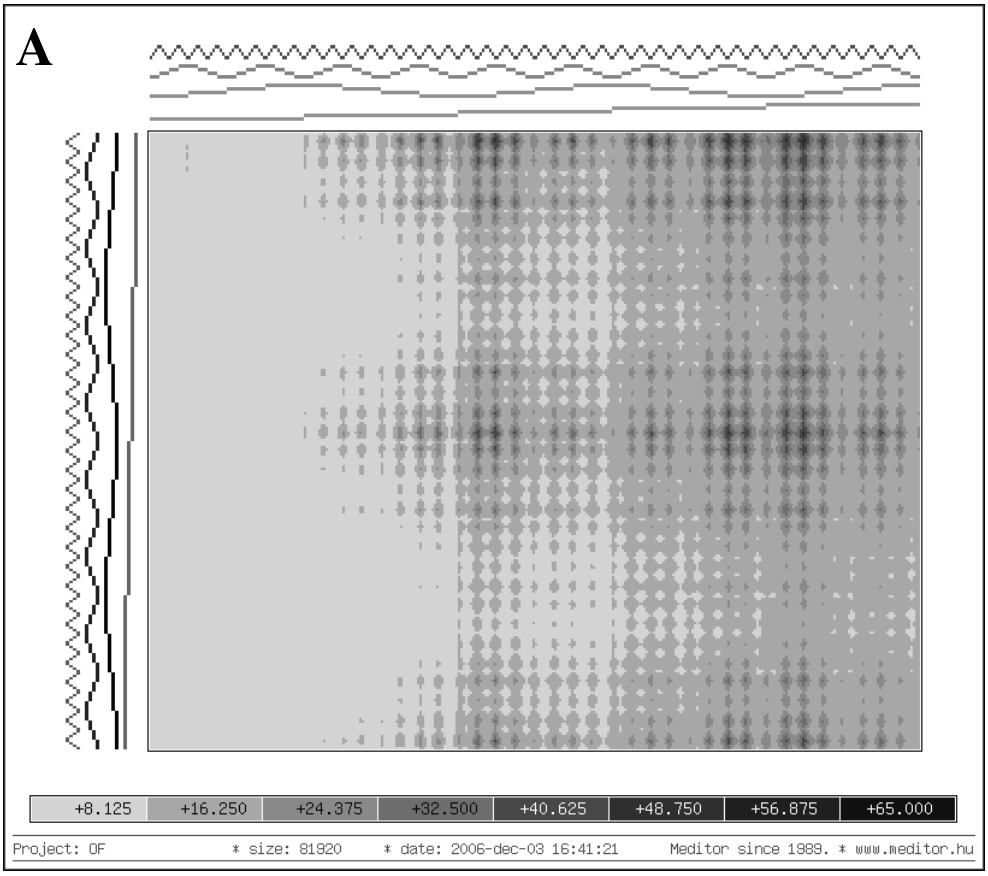
A
+8.125
+16.250
+24.375
+32.500
+40.625
+48.750
+56.875
+65.000
Project: OF
* size: 81920
* date: 2006-dec-03 16:41:21
Meditor since 1989. * www.meditor.hu

Fig. (7A).

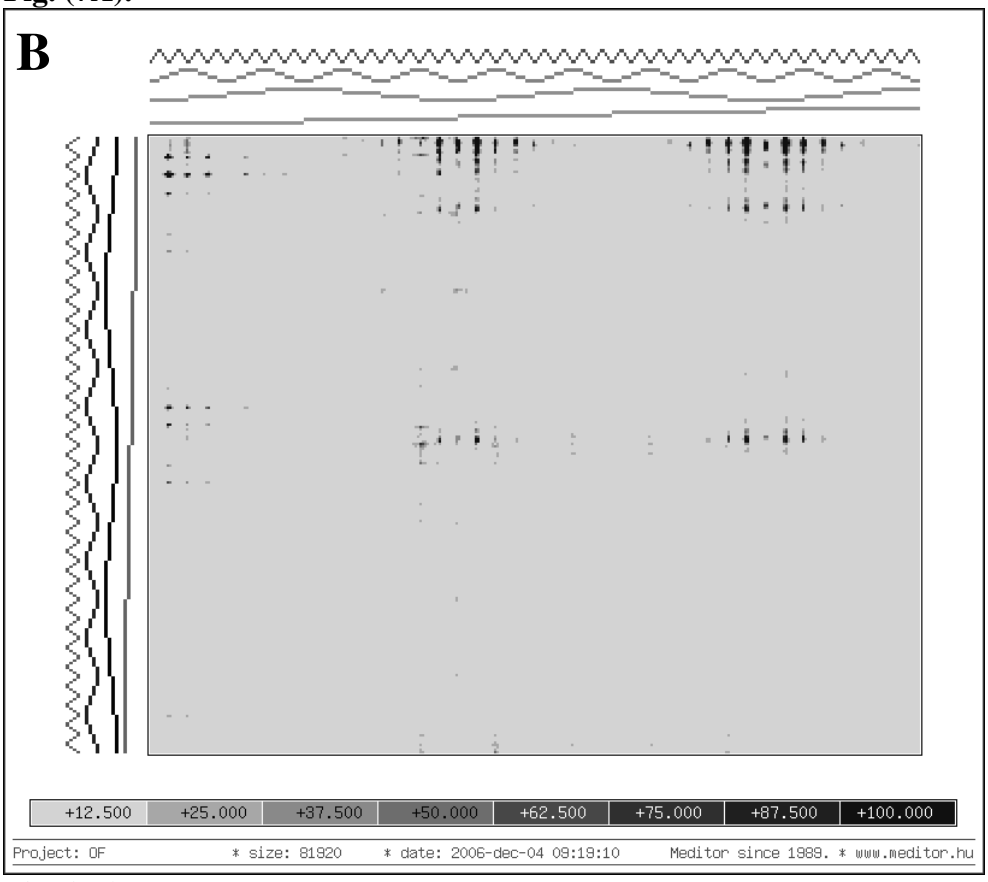
B
+12.500
+25.000
+37.500
+50.000
+62.500
+75.000
+87.500
+100.000
Project: OF
* size: 81920
* date: 2006-dec-04 09:19:10
Meditor since 1989. * www.meditor.hu

Fig. (7B).

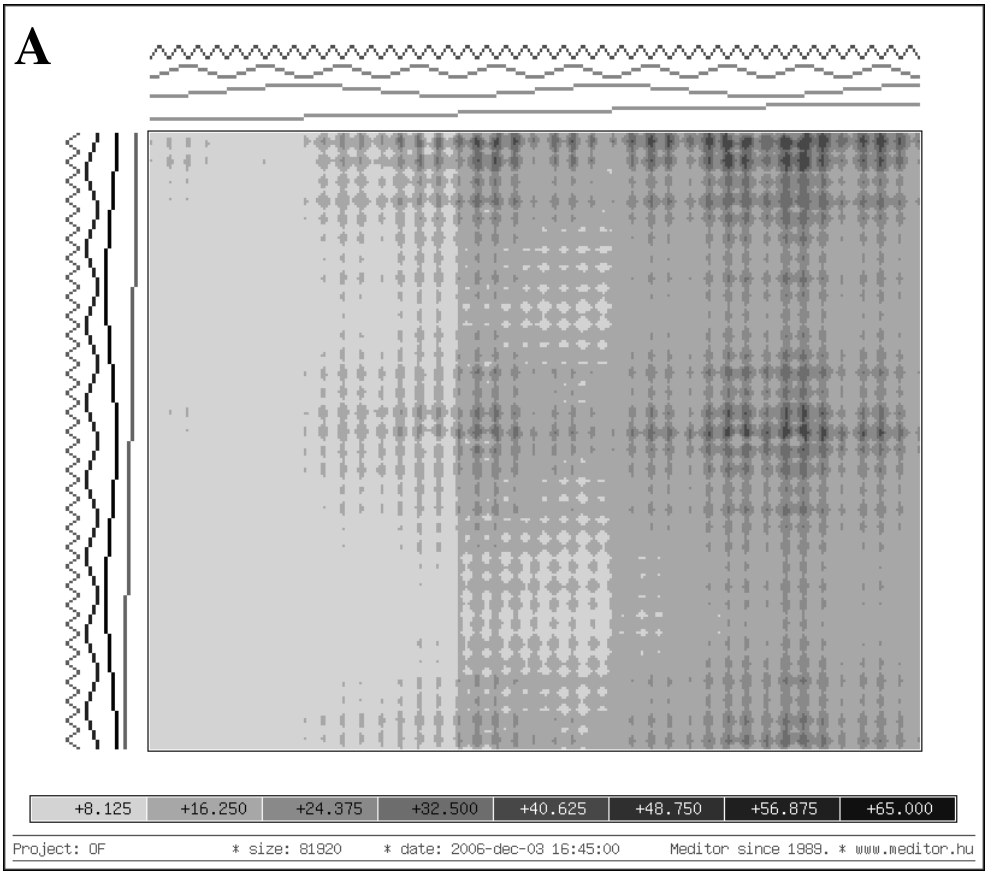
A
+8.125
+16.250
+24.375
+32.500
+40.625
+48.750
+56.875
+65.000
Project: OF
* size: 81920
* date: 2006-dec-03 16:45:00
Meditor since 1989. * www.meditor.hu

Fig. (8A).

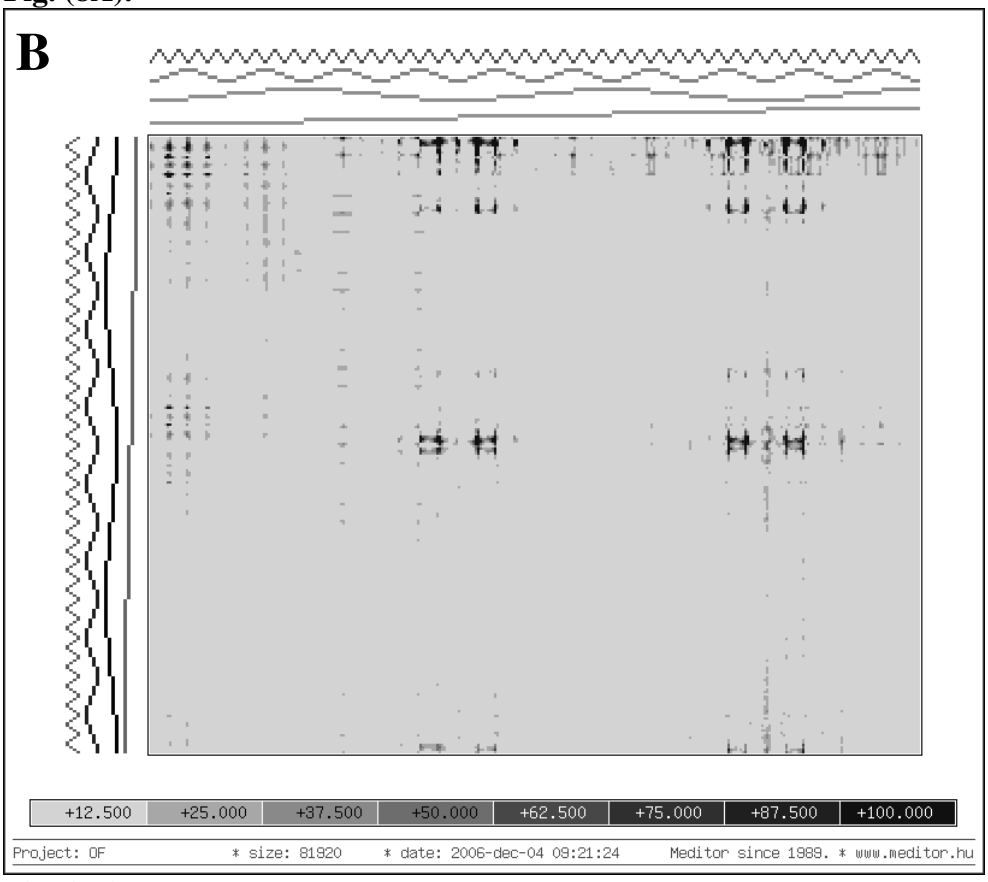
B
+12.500
+25.000
+37.500
+50.000
+62.500
+75.000
+87.500
+100.000
Project: OF
* size: 81920
* date: 2006-dec-04 09:21:24
Meditor since 1989. * www.meditor.hu

Fig. (8B).