An SOM-Hybrid Supervised Model for the Prediction of Underlying Physical Parameters from Near-Infrared Planetary Spectra*

Lili Zhang[1], Erzsébet Merényi[2], William M. Grundy[3], and Eliot F. Young[4]

[1] Rice University, Rice Quantum Institute and Department of Electrical & Computer Engineering MS-366, 6100 Main Street, Houston, TX 77005
llzhang@rice.edu
[2] Rice University, Department of Electrical & Computer Engineering MS-380, 6100 Main Street, Houston, TX 77005
erzsebet@rice.edu
[3] Lowell Observatory, 1400 W. Mars Hill Rd., Flagstaff AZ 86001
[4] Space Studies Department, Southwest Research Institute, Boulder, CO 80302

Abstract. Near-Infrared reflectance spectra of planets can be used to infer surface parameters, sometimes with relevance to recent geologic history. Accurate prediction of parameters (such as composition, temperature, grain size, crystalline state, and dilution of one species within another) is often difficult because parameters manifest subtle but significant details in noisy spectral observations, because diverse parameters may produce similar spectral signatures, and because of the high dimensionality of the feature vectors (spectra). These challenges are often unmet by traditional inference methods. We retrieve two underlying causes of the spectral shapes, temperature and grain size, with an SOM-hybrid supervised neural prediction model. We achieve 83.0±2.7% and 100.0±0.0% prediction accuracy for temperature and grain size, respectively. The key to these high accuracies is the exploitation of an interesting antagonistic relationship between the nature of the physical parameters, and the learning mode of the SOM in the neural model.

Keywords: Self-Organizing Map, parameter prediction, Near-Infrared spectra, New Horizons Space Mission, Pluto-Charon system.

1 Machine Learning for Parameter Prediction from Spectra, Motivated by the New Horizons Space Mission

1.1 Investigation of Surface Conditions from Near-Infrared Spectra

Near-infrared reflectance spectroscopy offers an extremely powerful remote probe of planetary surfaces. Numerous physical parameters, such as composition,

* This work was partially supported by grants NNG05GA63G and NNG05GA94G from the Applied Information Systems Research Program, NASA, Science Mission Directorate. Figures are in color, request color copy by email: llzhang@rice.edu, erzsebet@rice.edu

J.C. Príncipe and R. Miikkulainen (Eds.): WSOM 2009, LNCS 5629, pp. 362–371, 2009.

texture, and thermal state of surface materials, influence the observable reflectance, from which the parameters can potentially be retrieved. Icy outer Solar System surfaces present a natural application for such retrieval algorithms, since cryogenic ices such as H_2O, N_2, and CH_4 possess distinctive spectral contrasts which change in known ways as a function of temperature [1,2]. NASA's New Horizons spacecraft, which is en route to the Pluto system [3], will map the surfaces of Pluto, Charon, Nix, and Hydra at wavelengths from 1.25 to 2.5 microns with its infrared imaging spectrometer [4] in 2015. By extracting the surface parameters from these spectral maps, it will be possible to determine what processes are at work sculpting the exotic landforms New Horizons will discover.

However, the complexity of the measured spectra, their shapes, the often subtle changes in the spectral curves in response to relevant changes in the underlying causes (the implicit physical parameters of the surface), the interplay between the underlying causes, and the high dimensionality of the feature vectors (spectra), poses significant challenges for accurate retrieval of the parameters. Traditional approaches, based on iteratively inverting spectral mixing models [5] do not give entirely satisfactory results in real world applications [6]. Such techniques work very well for the relatively simple version of the problem posed here, but with the machine learning approach we propose we expect to be able to infer parameters also from complicated noisy real spectra. Specifically, we approach this challenge with a hybrid supervised neural architecture, which has a Self-Organizing Map (SOM) as its hidden layer. In this paper we focus on the inference of two surface parameters, grain size and temperature. We investigate inference capabilities of the neural model for a single material, crystalline water ice, which is common on Solar System surfaces. Results gained will be used in follow-up work to infer parameters from other types of ices (SO_2, CO, CO_2, NH_3) possibly occurring in the Pluto system and elsewhere in the Solar System.

1.2 Forward Training with Synthetic Spectra and Reverse Engineering from Real Spectra

The training of the neural machine requires a great number of spectra, which should span the meaningful ranges of the physical parameters with appropriate resolutions. However, real spectra collected from icy planetary surfaces are scarce, and not representative of the desired granularity of the prediction. To help this, synthetic or laboratory data have been used for model development in several areas, where real data are hard to obtain [7,8,9]. In our study the SOM-hybrid neural prediction model is trained with synthetic spectra, which are generated through radiative transfer code described in [10] based on the Hapke model, the most common way to represent the interaction of a solid surface with incident sunlight [5,11]. The synthetic spectra are generated on a grid of temperatures and grain sizes where the temperature ranges from 20 to 270 K with 2 K spacing, and the grain size takes 9 values logarithmically spaced from 0.0003 to 3.0 cm. This set of parameters encompasses the range of possible surface conditions of the icy Solar System bodies of our interest, at sufficient

resolution for scientific study. The spectral resolution, 230 band passes from 1 to 2.5 μm, is close to the resolution of the sensor used on the New Horizons spacecraft.

1.3 The SOM-Hybrid Supervised Architecture

The supervised neural architecture we use in this study is a fully connected feed-forward network with an SOM as the middle layer and an output layer connected to the SOM by the Widrow-Hoff rule [12]. A 230-band spectrum is taken as an input vector at every learning step. The learning consists of two stages. The first stage is unsupervised, in which the SOM layer captures the structure of the data manifold. The knowledge represented in the SOM is then utilized in the second stage, which is supervised training of the output layer. This construction generally helps achieve good prediction accuracy [13,14]. Its additional merits include ease and economy of training and handling of high-dimensional data, compared to other, more frequently used neural approaches. We use the Conscience variant [15] of the original Kohonen SOM [16]. It introduces a bias to achieve equal winning probabilities across all neural units thus producing more faithful *pdf* matching than the Kohonen SOM. Briefly, the weight vector $\mathbf{w}_i$ of neural unit i in the SOM lattice A of N neural units, is updated iteratively through a two-step procedure. First, a winner (or best matching unit, BMU) $\mathbf{w}_i$ is selected for a given input vector $\mathbf{x}$ such that with the bias b_j for neural unit j

$$\| \mathbf{w}_i - \mathbf{x} \|^2 - b_i \leq \| \mathbf{w}_j - \mathbf{x} \|^2 - b_j, \forall j \in A. \tag{1}$$

The bias b_j is computed from the winning frequency p_j, of neural unit j, as

$$b_j = \gamma(t) \times ((N \times p_j) - 1), \tag{2}$$

where γ is a parameter. Second, all weight vectors $\mathbf{w}_j$ are updated:

$$\mathbf{w}_j^{new} = \mathbf{w}_j^{old} + \alpha(t) h_{i,j}(t)(\mathbf{x} - \mathbf{w}_j^{old}). \tag{3}$$

Here, $h_{i,j}(t)$ is a neighborhood function, α is the learning rate. With the Conscience algorithm $h_{i,j}(t)$ can be fixed and of small size (e.g., the immediate neighbors in a diamond or square configuration), instead of a large neighborhood (e.g., Gaussian) that has to decrease with time.

2 Relationship of the Underlying Causes of Spectra as Seen from the SOM

The two physical parameters have different influences on the spectral shapes. As shown through two representative examples in Fig. 1, there is a substantial difference in the reflectance values of the spectra between two neighboring grain sizes, deepening of the absorption bands as well as shifting to lower values, in a linear fashion without crossing over. The increase in temperature from 30

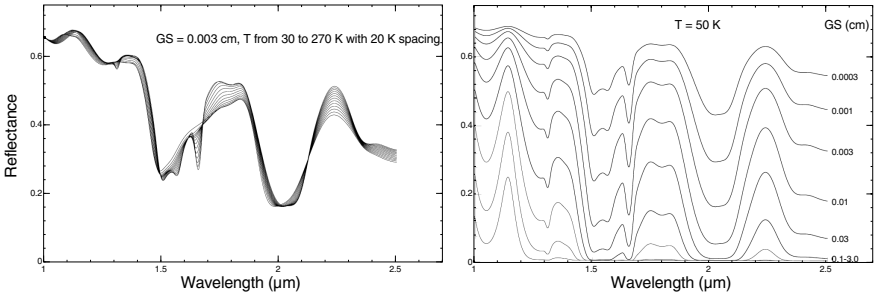

Fig. 1. Sample synthetic spectra of crystalline water ice. **Left:** variation in the spectral shape as a function of temperature, for one fixed grain size, 0.003 cm. **Right:** variation of the spectral shape as a function of grain size, at 50 K.

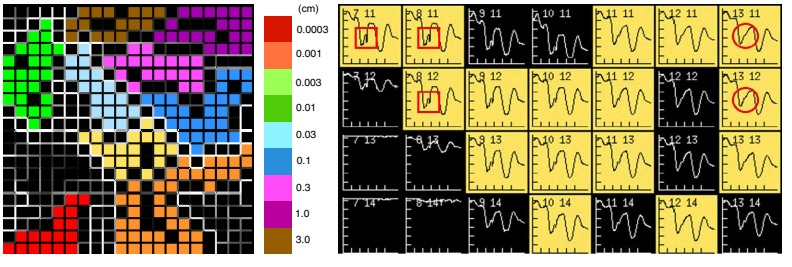

Fig. 2. Left: A 20×20 SOM trained with synthetic spectra of crystalline water ice. The colors represent the known grain size labels as keyed at right. The "fences" drawn between cells have grey scale intensities proportional to the Euclidean distance of the respective weights. White is large distance (dissimilarity). Black cells within double fences separating grain size groups indicate empty weights (no data mapped to them). Other black areas indicate weights representing spectra of other ices (N_2, CH_4 etc.). **Right:** An example of how spectra are organized within a grain size group, according to temperatures. Shown here are the learned weight vectors in the respective SOM cells, for the 0.003 cm (yellow) group. We can observe a continuous change in the spectral shapes from left to right, caused by increasing temperature. The red boxes and circles exemplify differences in absorption features at low and high temperatures, respectively.

to 270 K causes much smaller changes in reflectance but the direction of the change varies with wavelength. For example, the absorption at 1.65 μm gradually weakens with increasing temperature, while in the neighboring window of 1.7 – 1.9 μm the reflectances decrease with increasing temperature, resulting in crossovers. Fig. 1 suggests that the (Euclidean) distance between two grain size groups is larger at most wavelengths than the variations caused by temperature within that group. This dominance is reflected in the SOM after unsupervised learning by clearly separated clusters with respect to grain size (Fig. 2, left). Within each grain size cluster, the weight vectors show the continuous change in the spectral shapes caused by temperature, seen in Fig. 1. Fig. 2, right, illustrates

this for the 0.003 cm grain size (yellow) group. The weight vectors learned from spectra with low temperatures have a strong absorption at 1.65 μm (in red boxes). This feature gradually disappears for high temperatures (in red circles).

3 Conjoined Twin Machines

3.1 Two Modes of the SOM during Supervised Learning

The SOM can be used in two modes during supervised training. One is the widely used winner-take-all (WTA) mode, in which one SOM neuron fires (has an output signal of 1) in response to a given input vector, the rest send 0 to the weighted sums formed at the output layer. The other possibility is to divide the winning credit among k SOM nodes by assigning each an output value that is inversely proportional to the distance between its associated weight vector and the input vector (such that the credits add up to 1). This can be called "interpolating mode". For practical purposes k can be a number much lower than the number of neurons in the SOM, based on the assumption that the winner weight's Voronoi cell has a relatively low number of neighbor Voronoi cells in data space. For $k > 1$ we will use the term "interpolation on" or "interpolating mode", and use "interpolation off" or "non-interpolating mode" for $k = 1$.

For this data set, the Voronoi cells of all weight vectors of the SOM in Fig. 2, left, except for one, have at most 3 neighbors. We see this from the numbers of connections to Voronoi neighbors (pairs of BMUs and second BMUs formed by weights and their Voronoi neighbors) [17], shown in the order of the most to least connected, in Table 1. The connection strength (the number of data samples selecting a weight and its Voronoi neighbor as a pair of BMU and second BMU) between the one weight that has a fourth neighbor, and that fourth most connected neighbor is 1 (negligible). This justifies $k = 3$ for interpolating mode.

Table 1. Number of connections to Voronoi neighbors, from the most connected to the least connected, summed across all SOM weights

	Most connected	Second most connected	Third most connected	Fourth most connected
Number of connections	280	214	13	1

3.2 The Effect of SOM Modes on the Prediction Accuracies of Temperature and Grain Size

As observed in Section 2 the grain size dominance on the reflectance spectra causes clustering primarily by grain size in the SOM. Closer inspection reveals that, without exception, all spectra mapped to any weight vector within a grain size cluster have the same grain size label. This provides a good basis for perfect learning of grain size in non-interpolating mode, where only the BMU fires. In interpolating mode, each input spectrum stimulates the second and the third

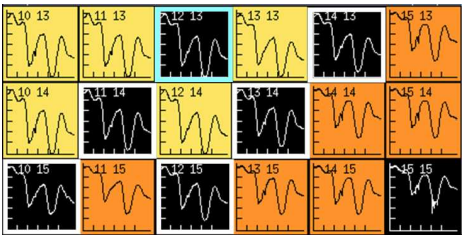

Fig. 3. The learned weight vectors shown in the respective SOM cells. The white boxes highlight weight vectors on the boundary between the yellow and orange groups, which represent 0.003 and 0.0001 cm grain sizes, respectively. The light blue box indicates an empty weight vector inside the yellow group.

BMUs, too, which may belong to the boundary area like the empty weight vectors in white boxes in Fig. 3. Their shapes are similar to both neighbors, which makes them candidates to be the second or third BMU for an input vector from either group. This introduces possible confusions in the supervised training. In contrast, the second and third BMUs may help refine the prediction of temperature. Since 126 spectra with different temperature parameters are forced to share approximately 25 – 30 SOM weights in a grain size cluster (Fig. 3, left), each weight forms an average (a mixture) of spectra, and each spectrum is likely to contribute to the mixture in several neighboring weights (smearing across neighbor weights) during training. None of these mixtures will match any specific temperature exactly, but a specific temperature may be reconstructed from several neighboring weight vectors by training the output weights to form their appropriate mixture. This includes empty weight vectors within any grain size group too, such as the one in the light blue box in Fig. 3.

The above discussion suggests conflicting preferences for SOM modes in the prediction of the two parameters. Supervised training results shown as correlations between predicted and true values in Fig. 4 confirm this. Since both physical parameters have large ranges, we quantify the prediction accuracies as

Table 2. The prediction accuracies of grain size (GS) and temperature (T) for two separate data sets, containing 9 and 81 grain sizes, respectively, with 20×20 and 40×40 SOMs, each in interpolating and non-interpolating modes. Results for the data with 9 GS are averages of 10 jack-knife runs. Results for the data with 81 GS are from a single run for reasons of time limitations. Further jack-knife runs are in progress.

		Data with 9 grain sizes		Data with 81 grain sizes	
SOM mode		Non-interpolating	Interpolating	Non-interpolating	Interpolating
20×20 SOM	GS	**100.0±0.0%**	76.4±4.4%	73.0%	**77.9%**
	T	76.2±2.6%	**83.0±2.7%**	32.9%	**53.7%**
40×40 SOM	GS	-	-	**97.7%**	54.3%
	T	-	-	61.2%	**77.3%**

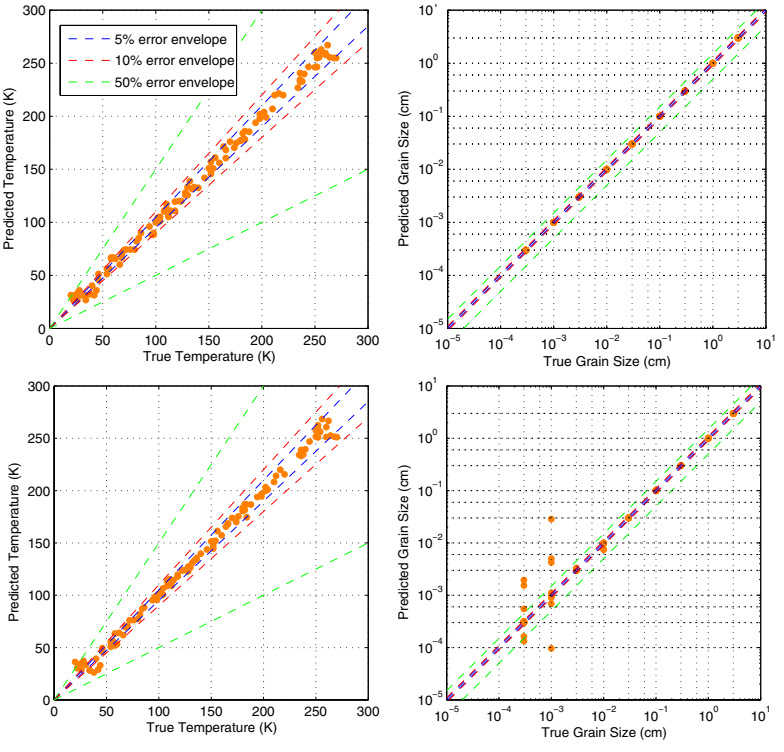

Fig. 4. Correlation of predicted (retrieved) and true values of temperature (**left block**) and grain size (**right block**). Data are shown as orange dots. **Top:** results obtained with the SOM in non-interpolating mode. **Bottom:** results obtained with the SOM in interpolating mode. The blue, red and green dashed lines indicate 5%, 10% and 50% error envelopes, respectively. Temperature has a smaller prediction error with interpolating mode. The prediction of grain size is better with non-interpolating mode.

the percentages of test data samples with less than 5% relative error. We achieve $100.0 \pm 0.0\%$ accuracy for grain size in non-interpolating mode, and $83.0 \pm 2.7\%$ for temperature in interpolating mode (Table 2, left block). Contrary to expectation, increasing the grid resolution of the grain size in generating training spectra does not help improve the prediction accuracy in interpolating mode, as shown in Table 2, right block. In fact, with a 20×20 SOM, the predictions of both temperature and grain size are significantly worse (or at best comparable), in both SOM modes, than results produced with 9 grain sizes. This is likely a result of the softening of boundaries between grain size groups due to an additional eight spectral curves between each two in Fig. 1, right. For this comparison we use an augmented training set generated with 81 grain sizes, equally spaced on a logarithmic scale in the same 0.0003 – 3.0 cm range as, and including, the training set with 9 grain sizes. Increasing the size of the SOM to 40×40 still does not improve the prediction of either parameter in interpolating mode. In

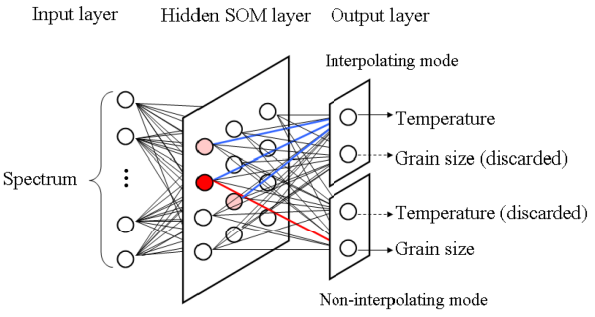

Fig. 5. Conceptual diagram of the Conjoined Twin Machines. One head of the Conjoined Twins works in non-interpolating mode, using the output from the BMU (red neuron in the SOM) to predict grain size. The other head, working in interpolating mode, uses, in addition, the second and third BMUs (pink) to predict temperature.

non-interpolating mode grain size prediction recovers to 97.9% because the space in the larger SOM allows the boundaries to become better defined again. However, it is unable to recover to 100% accuracy, because some of the boundary weight vectors still represent spectra with inhomogeneous grain size labels. On average, the 40×40 SOM allocates $\sim$ 19 weight vectors to each grain size group, which is $\sim$ 2/3 of the number allocated by the 20×20 SOM for 9 grain size groups. This is likely to be the cause of the unresolved boundaries.

We can conclude from the above that larger SOM size and more grain size samples do not get us closer to better overall prediction with a uniform interpolation scheme. While it is possible that with an even larger SOM we may be able to achieve the same accuracies as in Table 2, left block, the extra resources and time required make that solution undesirable for practical purposes. Instead, we can exploit the knowledge that the two parameters have opposing preferences for interpolation, and encode this duality into the learning machine to combine the advantages of the two SOM modes. This is the concept of the Conjoined Twin Machines (Fig. 5), which includes a shared SOM, containing the learned view of the manifold topology. Two identical "heads" both pull information from this shared SOM but each interprets it somewhat differently. One uses only the output of the BMU, and treats the rest of the SOM outputs as zeros thus not allowing them to influence the learning. This corresponds to non-interpolating mode and therefore will help best predict the grain size. While this grain size specialist head has a second output node identical to its twins', the prediction resulting from that node is discarded. Similarly, the second "head" specializes on temperature, by pulling the outputs of the first three BMUs into the weighted sums for training the output layer. This corresponds to interpolating mode with $k = 3$ and helps predict the temperature accurately while the grain size prediction is discarded. The final output of this machine is the grain size prediction from the first, and the temperature prediction from the second "head".

4 Conclusions and Future Work

This paper proposes an effective approach to predict two underlying physical parameters from near-infrared synthetic spectra with high accuracies. The architecture of the learning machine is in the form of Conjoined Twin SOM-hybrid supervised networks. We achieve 100.0±0.0% and 83.0±2.7% prediction accuracies for grain size and temperature, respectively. This means that for Charon, where temperatures in illuminated regions are likely to range up to 65 K, the neural model should be able predict temperatures with less than ~ 3 K error, for 80 – 86% of the measured spectra. This is valuable in resolving diurnal temperature changes on Charon, which provides the boundary condition to discover the processes in Charon's interior and in its atmosphere. To prepare for the real data returned by New Horizons, a noise sensitivity analysis is in progress to gauge the predictive power of learned models (learned with clean as well as noisy data) from spectra obtained in real circumstances.

Because of the observed interplay of the two parameters, in this study we could justify the choices of 1 and 3 for the interpolation granularity k for grain size and temperature, respectively. These choices are obviously data dependent, therefore cannot be automatically applied to other data without prior exploration of the data properties. Future work will include more — potentially interdependent — underlying parameters, in which case the Conjoined Twins can be extended to Conjoined Triplets, Quadruplets, or possibly to other tuplets, each with a different value of k. This will in turn motivate looking for automated ways to determine the optimal value of k for each machine, perhaps with meta-learning. Moving from a conceptual prototype to an integrated piece of software in the implementation of the "Conjoined Twins" is a short-term task, that will help with such future extensions.

References

1. Grundy, W.M., Schmitt, B.: The Temperature-Dependent Near-Infrared Absorption Spectrum of Hexagonal H_2O Ice. J. Geophys. Res. 103, 25809–25822 (1998)
2. Grundy, W.M., Schmitt, B., Quirico, E.: The Temperature Dependent Spectrum of Methane Ice I between 0. 65 and 5 Microns and Opportunities for Near-Infrared Remote Thermometry. Icarus 155, 486–496 (2002)
3. Young, L.A., et al.: New Horizons: Anticipated Scientific Investigations at the Pluto System. Space Science Reviews 140, 93–127 (2008)
4. Reuter, D.C., et al.: Ralph: A Visible/Infrared Imager for the New Horizons Pluto/Kuiper Belt Mission. Space Science Reviews 140, 129–154 (2008)
5. Hapke, B.: Theory of Reflectance and Emittance Spectroscopy. Cambridge University Press, New York (1993)
6. Grundy, W.M., et al.: Near-Infrared Spectra of Icy Outer Solar System Surfaces: Remote Determination of H_2O Ice Temperatures. Icarus 142, 536–549 (1999)
7. Gilmore, M.S., et al.: Effect of Mars Analogue Dust Deposition on The Automated Detection of Calcite in Visible/Near-Infrared Spectra. Icarus 172, 641–646 (2004)
8. English, E.C., Fricke, F.R.: The Interference Index and Its Prediction Using a Neural Network Analysis of Wind-tunnel Data. Journal of Wind Engineering and Industrial Aerodynamics 83, 567–575 (1999)

9. Gilbert, N., Terna, P.: How to Build and Use Agent-Based Models in Social Science. Mind & Society 1(1), 57–72 (2000)
10. Grundy, W.M.: Methane and Nitrogen Ices on Pluto and Triton: a Combined Laboratory and Telescope Investigation. PhD thesis, University of Arizona (1995)
11. Cruikshank, D.P., et al.: Ices on the Satellites of Jupiter, Saturn, and Uranus. In: Solar System Ices, pp. 579–606. Kluwer Academic Publishers, Dordrecht (1998)
12. Widrow, B., Smith, F.W.: Pattern-Recognizing Control Systems. In: Computer and Information Sciences (COINS) Symposium Proceedings, pp. 288–317. Spartan Books, Washington (1964)
13. Howell, E.S., Merényi, E., Lebofsky, L.A.: Classification of Asteroid Spectra Using a Neural Network. Jour. Geophys. Res. 99(E5), 10,847–10,865 (1994)
14. Rudd, L., Merényi, E.: Assessing Debris-Flow Potential by Using AVIRIS Imagery to Map Surface Materials and Stratigraphy in Cataract Canyon, Utah. In: Green, R. (ed.) Proc. 14th AVIRIS Earth Science and Applications Workshop, Pasadena, CA, May 24-27 (2005)
15. DeSieno, D.: Adding a Conscience to Competitive Learning. In: IEEE Intl Conference on Neural Networks, vol. 1, pp. 117–124 (1988)
16. Kohonen, T.: Self-Organizing Maps, 2nd edn. Springer, Heidelberg (1997)
17. Taşdemir, K., Merényi, E.: Exploiting the Data Topology in Visualizing and Clustering of Self-Organizing Maps. IEEE Trans. Neural Networks (in press, 2009)