Clustering and Classification based on Expert knowledge propagation using Probabilistic Self-Organizing Map: application to geophysics

Méziane YACOUB[1,2], Dominique FRAYSSINET[1,3], Fouad BADRAN[3,1], and Sylvie THIRIA[1]

[1] Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC) case 100, Université Paris 6, 4 place Jussieu 75252 Paris cedex 05 France
[2] Laboratoire d'Informatique de Paris Nord (LIPN), Inst. Galilée Université Paris 13, Av. J.B. Clément 93430 Villetaneuse France
[3] CEDRIC, Conservatoire National des Arts et Métiers, 292 rue Saint Martin 75003 Paris France

Abstract. In the present paper we describe a complete methodology to cluster and classify data using Probabilistic Self-Organizing Map (PRSOM). The PRSOM map gives an accurate estimation of the density probabity function of the data, an adapted hierarchical clustering allows to take into account an extra knowledge given by an expert. We present two actual applications of the method taken in the domain of geophysics.

1 Introduction

The Self-Organizing Map (SOM) as introduced by Kohonen (Kohonen (1984)), has been widely used for visualizing and clustering high-dimensional patterns. In that aspect, SOM is close to non linear projection methods such as Multi-Dimensional Scaling (MDS). Combined with cluster analysis as proposed in Bock (1997), MDS allows to obtain an optimal fit between a classification and its representation in a low-dimensional euclidian space. Visualization, as proposed in SOM methods uses a deformable discrete lattice to translate data similarities into spatial relationships. A large variety of related algorithms have been derived from the first SOM model (Oja and Kaski (1999)) which differ from one another but share the same idea to introduce a topological order between the different clusters. In the following we introduce a SOM algorithm, the Probabilistic Self Organizing Map (PRSOM) (Anouar et al. (1997)) which uses a probabilistic formalism. This algorithm approximates the density distribution of the data with a mixture of normal distribution. At the end of the learning phase, the probability density of the data estimated by the PRSOM map can be widely used to provide an accurate classifier. The first section of the paper is devoted to the PRSOM model, the second and the third sections show how to transform the probabilistic map using extra knowledge in order to obtain an accurate classifier. The efficiency of the approach is illustrated in two actual geophysical problems (the lithologic facies reconstruction and the classification of Top Of the Atmosphere (TOA) spectra of satellite ocean color measurements).

2 Probabilistic Self-Organizing Map (PRSOM)

As the standard Self Organizing Map (SOM) (Kohonen (1984)), PRSOM consists of a discrete set $(\mathcal{C})$ of formal neurons. Its map has a discrete topology defined by an undirect graph; usually it is a regular grid in one or two dimensions. For each pair of neurons (c,r) on the map, the distance $\delta(c,r)$ is defined as being the shortest path between c and r on the graph. For each neuron c, this discrete distance allows to define a neighborhood of order d: $V_c(d) = \{r \in \mathcal{C}/\delta(c,r) \leq d\}$. In the following, we introduce a Kernel positive function K ($\lim_{|x|\to\infty} K(x) = 0$) and its associated family K_T parametrized by T :

$$K_T(d) = [1/T]K(d/T) \tag{1}$$

The parameter T allows to control the neighborhood size.

Let $\mathcal{D}$ be the data space $(\mathcal{D} \subset \mathcal{R}^n)$ and $\mathcal{A} = \{\mathbf{z}_i; i = 1, \ldots, N\}$ the training data set $(\mathcal{A} \subset \mathcal{D})$. As the standard SOM algorithm, PRSOM defines a mapping from $(\mathcal{C})$ to $\mathcal{D}$ where a neuron c is associated to its referent vector $\mathbf{W}_c$ in $\mathcal{D}$. At the end of the learning algorithm two neighbour neurons on the map have close referent vectors in the euclidian space $(\mathcal{R}^n)$.

In contrast to SOM, PRSOM is a probabilistic model which associates to each neuron c a spherical Gaussian density function f_c. This density function is defined by its mean (referent vector) which is a n-dimensional vector, $\mathbf{W}_c = (w_c^1, w_c^2, \ldots, w_c^n)$ and its covariance matrix which is the diagonal matrix defined by $\Sigma_c = \sigma_c^2\mathbf{I}$. We denote by $\mathcal{W} = \{\mathbf{W}_c; c \in \mathcal{C}\}$ and $\sigma = \{\sigma_c; c \in \mathcal{C}\}$ the two sets of parameters defining the PRSOM model.

PRSOM allows us to approximate the density distribution of the data using a mixture of normal densities. In this probabilistic formalism, the classical map C is duplicated into two similar maps $\mathcal{C}_1$ and $\mathcal{C}_2$ provided with the same topology as $\mathcal{C}$. It is assumed that the model is a folded Markov chain (Luttrel (1994)), so for every input data $\mathbf{z} \in \mathcal{D}$ and every pair of neurons $(c_1, c_2) \in \mathcal{C}_1 \times \mathcal{C}_2$:

$$p(c_2/\mathbf{z}, c_1) = p(c_2/c_1) \quad and \quad p(\mathbf{z}/c_1, c_2) = p(\mathbf{z}/c_1) \tag{2}$$

It is thus possible to compute the probability of any pattern $\mathbf{z}$

$$p(\mathbf{z}) = \sum_{c_2} p(c_2)p_{c_2}(\mathbf{z}), \quad where \tag{3}$$

$$p_{c_2}(\mathbf{z}) = p(\mathbf{z}/c_2) = \sum_{c_1} p(c_1/c_2)p(\mathbf{z}/c_1) \tag{4}$$

The probability density $p_{c_2}(\mathbf{z})$ is completely defined from the network, giving the conditional probability $p(c_1/c_2)$ on the map and the conditional probability $p(\mathbf{z}/c_1)$ on the data. As we mentioned before, $p(\mathbf{z}/c_1)$ is a normal distribution: $p(\mathbf{z}/c_1) = f_{c_1}(\mathbf{z}, \mathbf{W}_{c_1}, \sigma_{c_1})$. If we assume that:

$$p(c_1/c_2) = [1/T_{c_2}]K_T(\delta(c_1, c_2)), \qquad where \qquad T_{c_2} = \sum_r K_T(\delta(c_2, r)), \quad (5)$$

The a posteriori density (??) can be expressed with respect to the normal distributions of the neurons:

$$p_{c_2}(\mathbf{z}) = [1/T_{c_2}] \sum_{r \in C_1} K(\delta(c_2, r)) f_r(\mathbf{z}, \mathbf{W}_r, \sigma_r) \quad (6)$$

Equation (??) shows that the data density $p(\mathbf{z})$ is a mixture of local mixtures of normal densities $f_{c_1}(\mathbf{z}, \mathbf{W}_{c_1}, \sigma_{c_1})$ whose mean vector and standard deviation are the parameters of the PRSOM model and have to be estimated from the learning data set $\mathcal{A}$ using the learning algorithm of PRSOM. The learning algorithm of PRSOM maximizes the likelihood of the learning set $\mathcal{A}$. It is a dynamic cluster method (Diday and Simon (1976)) operating in two steps which are run iteratively until convergence :

- **The assignment step** assigns each observation $\mathbf{z}$ to one cell c_2 using the assignment function χ (relation ??). This step gives rise to a partition of the data space $\mathcal{D}$, each observation $\mathbf{z}_i$ being assigned to the most likely cell according to the density p_{c_2} :

$$\chi(\mathbf{z}) = \arg\max_{c_2} p_{c_2}(\mathbf{z}) \quad (7)$$

- **The minimization step** assumes that the observations of the learning set $\mathcal{A}$ are independent and maximizes the conditional likelihood of the training data set $\mathcal{A}$ with respect to the parameters $(\mathcal{W}, \sigma)$ and under the hypothesis that each $\mathbf{z}_i$ is generated by the $p_{\chi(\mathbf{z}_i)}$ distribution.

$$p(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N/\mathcal{W}, \sigma, \chi) = \prod_{i=1}^{N} p_{\chi(\mathbf{z}_i)}(\mathbf{z}_i) \quad (8)$$

The minimization step minimizes the conditional log-likelihood function $E(\mathcal{W}, \sigma/\chi)$ according to $\mathcal{W}, \sigma$ and χ

$$E(\mathcal{W}, \sigma/\chi) = \sum_{i=1}^{N} -Ln[\sum_{r \in C} K(\delta(\chi(\mathbf{z}_i), r)) f_r(\mathbf{z}_i, \mathbf{W}_r, \sigma_r)] \quad (9)$$

The parameters $\mathcal{W}$ and σ are updated by setting the derivatives of $E(\mathcal{W}^k, \sigma^k/\chi^k)$ to zero (χ^k,$\mathcal{W}^k$ and σ^k denote the parameters at iteration k). To solve this equation we use, as in Duda (1973), an iterative approximation which assumes that, for the k^{th} iteration, the initial estimates of the parameters are sufficiently close to the true values. This leads to the following formulas:

$$\mathbf{W}_r^k = \frac{\sum_{i=1}^{N} \mathbf{z}_i K(\delta(r, \chi^{k-1}(\mathbf{z}_i))) \frac{f_r(\mathbf{z}_i, \mathbf{W}_r^{k-1}, \sigma_r^{k-1})}{P_{\chi^{k-1}(\mathbf{z}_i)}(\mathbf{z}_i)}}{\sum_{i=1}^{N} K(\delta(r, \chi^{k-1}(\mathbf{z}_i))) \frac{f_r(\mathbf{z}_i, \mathbf{W}_r^{k-1}, \sigma_r^{k-1})}{P_{\chi^{k-1}(\mathbf{z}_i)}(\mathbf{z}_i)}} \tag{10}$$

$$(\sigma_r^k)^2 = \frac{\sum_{i=1}^{N} ||W_r^{k-1} - \mathbf{z}_i||^2 K(\delta(r, \chi^{k-1}(\mathbf{z}_i))) \frac{f_r(\mathbf{z}_i, \mathbf{W}_r^{k-1}, \sigma_r^{k-1})}{P_{\chi^{k-1}(\mathbf{z}_i)}(\mathbf{z}_i)}}{n \sum_{i=1}^{N} K(\delta(r, \chi^{k-1}(\mathbf{z}_i))) \frac{f_r(\mathbf{z}_i, \mathbf{W}_r^{k-1}, \sigma_r^{k-1})}{P_{\chi^{k-1}(\mathbf{z}_i)}(\mathbf{z}_i)}} \tag{11}$$

PRSOM learning algorithm

Initialization phase: k=0 the initial parameters $\mathcal{W}^0$ are set using the SOM algorithm. χ^0 is defined according to equation (**??**) and σ^0 is computed according to (**??**). The maximum number of iterations (N_{iter}) is chosen.

Iterative step k $\mathcal{W}^{k-1}$ and σ^{k-1} are known

- Minimization step : update of the new parameters $\mathcal{W}^k$ and σ^k according to Eq.**??** and **??**.
- Assignment step : Update of the new assignment function χ^k associated to $\mathcal{W}^k$ and σ^k according to Eq. **??**

Repeat the iteration step until $k > N_{iter}$.

As in the Kohonen algorithm, PRSOM makes use of a neighborhood sytem of which the size, controlled by T, decreases as learning proceeds. At the end of the learning phase, the map provides the topological order; the partition associated with the map is defined by the latest assignment function $\chi^{N_{iter}}$ as defined in (**??**) and is based on probability considerations. Therefore the partitions provided by PRSOM are different from those provided by SOM which use euclidian distance. The estimation of the density function gives an extra information which will be of interest later for classification purpose. At the end of the learning phase, $\mathcal{D}$ is divided in M subsets : each neuron c of the map represents a particular subset $\mathcal{P}_c = \{\mathbf{z}/\chi^{N_{iter}}(\mathbf{z}) = c\}$.

3 PRSOM-Classifier

The map provided by PRSOM (as the one provided by SOM) can be used in classification tasks, combining supervised and unsupervised learning. This can be done by labelling each neuron c of the map using labels. If the data are distributed in S classes, denoted by $\mathcal{L} = \{l_i, i = 1, \ldots, S\}$, each pattern $\mathbf{z}$ is assigned to a neuron $c = \chi(\mathbf{z})$ of the map and takes the label l_c of the neuron c. So the labelled map can be used as a "hard" classifier. The problem is thus to distribute the labels amoung the neurons of the map according to some expert knowledge or to some contiguity characteristics which have to be defined. If there are a large amount of labelled data, they allow us to determine a consistent classifier. Each neuron of the map C being labelled using the majority vote rule. If we have a limited number of labelled patterns, the majority vote does not provide a label to each neuron and does not ensure a valid labelling. The problem is thus to propagate the valid labels in order to obtain a coherent labelled map and a consistent classifier.

At the end of the labelling process the set of neurons $\mathcal{C}_i$ which have the same label l_i correspond to different density functions which approximate the probability density function of classe l_i.

As PRSOM is a probabilistic model, a new pattern $\mathbf{z}$ can be classified with PRSOM computing the a posteriori probality of each class l_i:

$$p(l_i/\mathbf{z}) = \frac{\sum\limits_{c_2 \in \mathcal{C}_i} n_{c_2} p_{c_2}(\mathbf{z})}{\sum\limits_{c_2 \in \mathcal{C}} n_{c_2} p_{c_2}(\mathbf{z})} \tag{12}$$

where n_{c_2} represents the number of patterns in the subset $\mathcal{P}_{c_2}$. In this formula we estimate $p(c_2)$ (equation ??) by $\frac{n_{c_2}}{N}$.

The knowledge of the S a-posteriori probabilities allows us to define a "soft" classifier: for a given pattern $\mathbf{z}$, it is possible to propose different classes with their related probabilities. These a-posteriori probabilities vary with the labelling of the map, their consistency depending on the accuracy of that map, it is related not only to the amount of expert knowledge but also to the ordering of the topological map. Some methodology has to be defined in order to improve this first classifier when a small amount of knowledge is available. In the following section we present the approach we use for the two applications, presented in section (??).

4 HC-PRSOM : Hierarchical Clustering of PRSOM

To propagate the labels of the PRSOM map when necessary, two cases have to be considered according to the number of neurons labelled at the precedent stage (see section ??):

- If this number is zero, which means that we don't have any expert knowledge, a clustering of the neurons allows to propose classes which can be labelled by an expert according to external criteria.
- When the number is less than M, an adapted clustering allows to spread the knowledge of the labelled neurons on the remaining ones.

At this point, we choose the widely used hierarchical clustering (HC), to cluster the neurons of the map (see Murtagh in (Thiria et al. (1997))). HC algorithm can be summarized as follows :

1. Assign each pattern to a unique cluster.
2. Find the two closest clusters according to some dissimilarity and agglomerate them to form a new cluster
3. Return to step (2) until all patterns fall into one cluster.

Various algorithms can be constructed depending on the dissimilarity used in step (2). As shown in equation (**??**), the partition provided by PRSOM is computed according to the different variances of the different subsets. So the use of the Ward index which minimizes the decrease of inter class variance, takes advantage of the qualities of the partition provided by PRSOM : HC used on PRSOM (HC-PRSOM) with the Ward index provides clusters which take into account an accurate estimate of the variances.

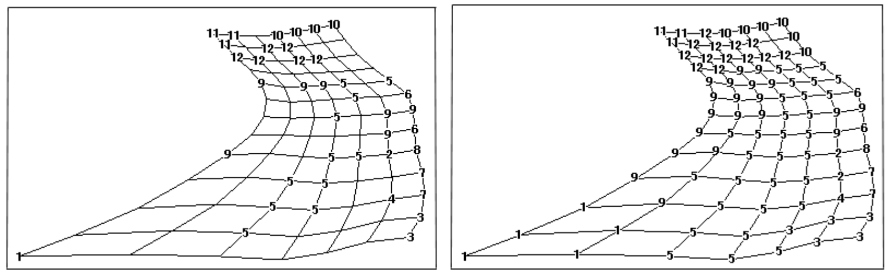
If we are in the first case where there is no labelled data, HC-PRSOM provides a classification whose labels have to be chosen by an expert. This can be done afterwards by a careful examination of the S subclasses of the partition.

If the knowledge is scarce (few neurons have been labelled), HC can be adapted in order to propagate the label of the neurons. In that case, the three steps of HC are modified as follows, giving rise to an extended HC-PRSOM (εHC-PRSOM) classification :

1. Assign each neuron to a unique cluster.
2. Find the two closest clusters according to the Ward index.
   - If they have the same label, they give rise to a cluster with the same label.
   - If just one cluster has a label, the agglomerate cluster and all its neurons take this label.
   - If none of the two clusters are labelled, no label is added.
   - If the two clusters have distinct labels, each one keeps its own label, the clustering does not behave and the algorithm looks for the next two closest clusters.
3. Return to step (2) until all the neurons have received a label.

At the end of εHC-PRSOM, a partition of S sets of neurons representing the S different classes is provided.

In the next section, in order to illustrate PRSOM behaviour, we present two real applications taken in the field of geophysics.



(a) (b)

Fig. 1. This figure shows the Sammon projection, in a two dimensional space, of the referent vectors with the topological order induced by the map. (a) Shows the partially labelled map after introducing the expert knowledge, (b) shows the labelled map after propagating the expert knowledge.

5 Application

5.1 Lithologic facies reconstruction

In order to investigate the subsoilís composition, the geologists take down sensors for measuring physical properties of the rocks into drilling hole. We call logs, the result of these measures (for more details on loggin tools, see (Rabaute (1999)). In the present application, we consider five logs (RHOB, GR, PEF, NPHI, and DT) taken from two different holes : $A1$ and $A2$, drilled in the sea floor; RHOB is the rock density, Gr measures the natural radioactivity, PEF is the photoelectric log, NPHI is the neutron porosity and DT is an acoustic measure. So each data $\mathbf{z}$ used for the classification task is a five-dimensional vector made of the five measures taken from one point in the drilling hole.

In the following, we use of the full set of 1717 measurements we have in order to train the different algorithms : card($\mathcal{A}$)=1717. The set $\mathcal{A}$ has been used to train a $7 \times 13 = 91$ neuron map using first SOM, then PRSOM. The first map provided by SOM is improved by PRSOM, which enables us to estimate the mixture of normal densities . At this point of the method, $\mathcal{D}$ is partitioned into 91 classes which correspond to the 7×13 neurons.

Geologists chose to drill this part of the sea-floor, because they have a good knowledge about the nature of the subsoil in this place, and they suspect they may find some oilfields. They want to put together the logís observations into the 12 geologic classes existing in the drilling holes $A1$ and $A2$. Moreover, some cores have been extracted in several parts of the holes, so the geologists certainly know which kind of rocks exists for some given depth. We introduce this expert knowledge into the topologic map, thanks to the majority vote. We obtain a map which is partially labelled as shown on figure ??(a). Then we cluster the 91 neurons into 12 classes, by using the new εHC-PRSOM classification propagating the labels all over the map. At this stage, we have

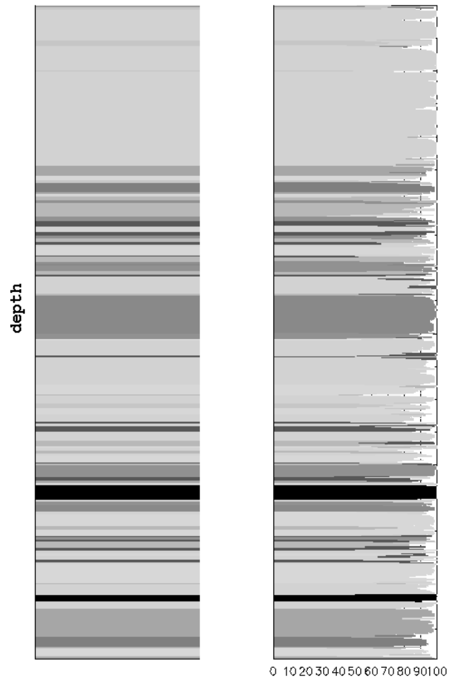


Fig. 2. This figure shows the reconstructed lithologic facies obtained for the drilling hole $A2$ at each depth of the hole, each geologic class is represented by a dedicated grey level. (a) Represents when using as label the one "hard classification" provided by the assignement functio χ. (b) Represents the a-posteriori probabilities computed using formula **??** ("soft" classification), each depth is represented using two or three different labels according to the a-posteriori probability. In this case the x-axis represents the probability, facies with probability less than 0.15 are not represented.

a wholly labelled map, shown on figure **??**(b). We can see in figure **??**(b) that the topology of the map reflects the similarity between the different geological classes. The coal, class number 1, is very different from the other rocks. It stands at one extremity of the map. The top of the map is taken up by the classes 10, 11 and 12 which are all coarse and medium-grained sandstones. The classes 2, 3 and 4 are different shales, which are all to be found in the bottom right-hand corner. In the middle of the map, we find the paleosols, class number 5. The paleosols consist of shales which have been enriched with a diversity of components, so they are proximate to every other class. The micaceaous sandstones are separated into two groups, one part near the coarse grained sandstone, and one part near the micaceous silts, class number 6. Micaceous sandstones can be slipped into two different classes. The geologist has pointed out that micaceous silts (class number 6) and very fine grained sand (class number 8) can very close properties : they are next to each other on the map. Reporting the class number of each neuron to the observations which are associated to this neuron allows us to decompose a lithologic facies in 12 classes. Figure **??**(a) shows the facies obtained from the drilling hole $A2$.

Thanks to the PRSOM algorithm, each neuron has a normal density function, so we can introduce probability criterium in the classification (formula **??**). These a-posteriori probabilities are computed using the map labelled using εHC-PRSOM. It is worth noticing that the different classes are very dependent on the majority vote. We thus have determined the probability that an observation belongs to each of the 12 classes. Plotting the probabilities and the classification reflects the reality of the drilling holes better. The figure **??**(b) shows the selection of the classe of best probability for each observation, in the well $A2$. For each point of the holl, we only draw the classes for which the rocks belongs to with a probability above 15%. That is why we do not cover the 100% of probability all along the well.

5.2 Satellite ocean color classification

Satellite ocean color sensors which are now operational or under preparation, measure ocean reflectance, allowing us to a quantitative assessment of geophysical parameters (e.g., chlorophyll concentration).

The major problem for ocean color remote sensing processing is that interactions of the solar light with aerosols in the atmosphere and water constituents are responsible for the observed spectrum of Top Of the Atmosphere (TOA) reflectances for a cloud-free pixel. The critical issue is thus to remove the aerosol contribution to the measured signals in order to retrieve the actual spectrum of marine reflectance which is directly related to the water constituent content. Ocean color is determined by the interactions of the solar light with substances or particles present in the water, and provides useful indications on the composition of water.

In case of cloud or earth pixels, there is no information on ocean color, so the first step is to remove them. In the following we used PRSOM and HC-PRSOM in order to clean the set of TOA reflectances.

Data We used SeaWiFS data product. The SeaWiFS[1] on board the SeaStar satellite is a color-sensitive optical sensor used to observe color variations in of the ocean. It contains 8 spectral bands in the visible and near-infrared wavelengths [2]. SeaWiFS ocean color data is available in several different types. We used level 1 GAC data : it consists of raw radiances measured at the top of the atmosphere.

[1] SeaWiFS Project Home Page http://seawifs.gsfc.nasa.gov/SEAWIFS.html

[2] 412nm, 443nm, 490nm, 510nm, 555nm, 670nm, 765nm and 865nm

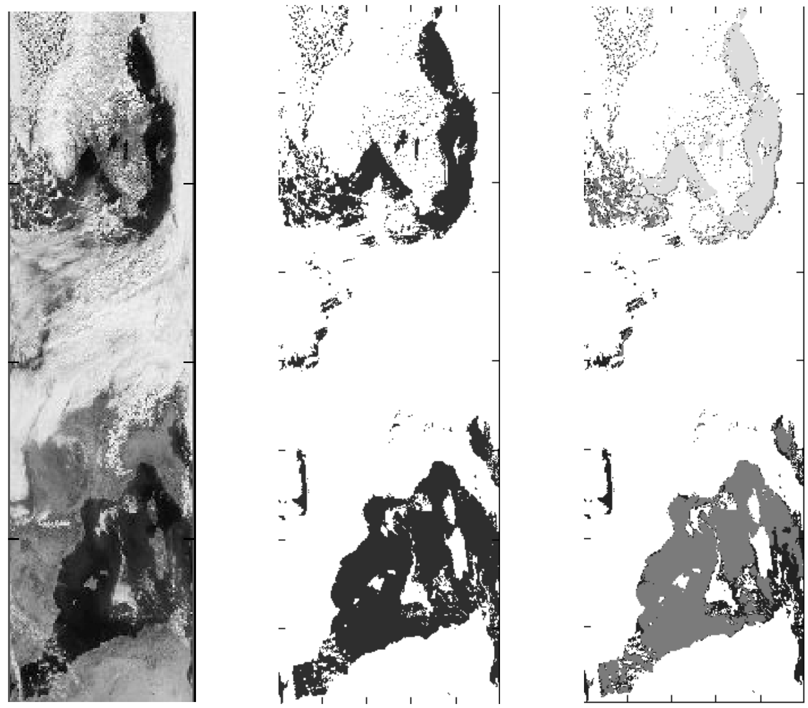
We applied PRSOM algorithm to the image shown in figure ??(a). It represents the Baltic sea, which is rich in turbid case-2 water, and the West Mediterranean sea. The image (194680 pixels) was taken on February 1998.

Experimental results In the first step we are interested in separating water pixels from land and clouds pixels, and in the second step to separate different area of water on a selected number of classes, each class being represented by its mean apparent optical properties.
In order to separate water pixels from land and cloud pixels, we used only the spectral bands $670nm$, $765nm$, and $865nm$, in which the signal coming from the ocean is nul or negligible. We trained a two-dimensional map of size 10×10 with PRSOM. Then we used HC to cluster the referent vectors of the PRSOM map and cut the dendrogram in order to obtain 2 clusters. The two subsets represented by each cluster are shown in figure ??(b). The white region corresponds to the land or clouds pixels, and the black one represents the water pixels. We compared our results to those proposed by the SeaWiFS algorithm. Table **??** displays the distribution of the pixels between the two classes water and land or clouds according to the classification proposed by SeaWiFS algorithm and that obtained when using the combination of PRSOM and HC (HC-PRSOM). It may be noticed that 96.28% pixels are classified in the same way by SeaWiFS algorithm and HC-PRSOM. Furthermore, all the pixels which are classified as land or clouds by SeaWiFS are classified in the same way by HC-PRSOM, 3.72% pixels are classified water by HC-PRSOM and land or clouds by SeaWiFS. This difference can be explained as follows : SeaWiFS separates water from land and clouds (light or thich) while HC-PRSOM separates water from land and thick clouds only.

		SeaWiFS results	
		land or Clouds	**visible water**
HC-PRSOM results	**land or Clouds**	**149791**	**0**
	visible water	**7235**	**37654**

Table 1. Distribution of the pixels between the two classes land or clouds, and water

To separate different areas of water on a selected number of classes, we applied HC-PRSOM only to the water pixel (the black region in figure ??(b)) using in this case the 8 spectral bands. As before, we used HC-PRSOM on a two-dimensional map of size 10×10. The three regions obtained when we cut the dendrogram in order to obtain 3 clusters are shown in figure ??(c). According to the ocean expert, light grey region corresponds to the water in the Baltic sea, which contains a high level of yellow substances. The grey region corresponds to clear water, usually present in the Mediterranean sea, and the



(a) (b) (c)

Fig. 3. (a) : the studied region. (b) : the two classes obtained when using HC-PRSOM to separate water pixels from land and clouds pixels. (c) : the regions obtained when separating water pixels in three classes. In (b) the black region corresponds to water pixels and the white one corresponds to land or clouds pixels. In (c), the light grey region corresponds to the water in the Baltic sea, while the grey region correponds to clear water, and the black region corresponds to a thin cloud cover over the sea.

black region corresponds to a thin cloud cover over the sea. The three classes obtained present a significant coherence with geophysical knowledge of the ocean expert. Notice that HC allows to obtain up to 100 classes (since we used a topological map of size $10X10$), the other classes actually being under interpretation. We are exploring whether we can refine the classification in water type and find other, previously unidentified classes.

6 Conclusion

In the present paper, we have described a complete methodology to cluster and classify data using Probabilistic Self-Organizing Map. The PRSOM algorithm gives an accurate estimation of the density probabity function of the data, an adapted hierarchical clustering allows us to take into account an extra knowledge given by an expert. The two actual applications, presented here, show that the approach is efficient. In both cases, using or not an extra knowledge expertise, provide classes which present strong spatial patterns in good agreement with the reality. A lot of improvements can be done for better taking into account the probabilitic aspect of the PRSOM algorithm. These improvements can be introduced at the labeling as well as at the hierachical clustering stage.

Acknowledgements

Part of the lithologic application was made in collaboration with Louis Briqueu from University of Montpellier II and Stéphanie Gottlib-Zeh from ELF Exploration Production (Pau, France). Part of the Satellite ocean color classification was made in collaboration with Michel Crepon, Jean-Michel André, Lidwine Gross and Carlos Mejia from LODYC, University of Paris 6, France. The authors would like to acknowledge them for providing us the databases and making useful comments about the results.

References

ANOUAR F., BADRAN F. and THIRIA S., (1997) : Self Organized Map, A Probabilistic Approach. proceedings of the Workshop on Self-Organized Maps. Helsinki University of Technology, Espoo, Finland, June 4-6.

BOCK, H.H. (1997) : Simultaneous visualization and clustering methods as an alternative to Kohonen maps. In: G. Della Riccia, R. Kruse, H.-J. Lenz (eds.): Learning, networks and statistics, CISM Courses and Lectures no. 382. Springer, Wien - New York, 67-85.

DIDAY E. and SIMON J.C. (1976) : Clustering Analysis. In Digittal Pattern Recognition, Edited by K.S.FU. Springer-Verlag

DUDA R.O., and HART P.E. (1973) : Pattern classification and scene analysis. John Wiley & Sons, Inc., New York.

RABAUTE A. (1999) : Obtenir une représentation en continu de la lithologie et de la minéralogie. Exemples d'applications du traitement statistique de données de diagraphie aux structures sédimentaires en régime de convergence de plaques (Leg ODP 134, 156 et 160). Thèse de doctorat, Université de Montpellier II. Mémoires géosciences - montpellier.

KOHONEN T. (1984) : Self organization and associative memory. Springer Series in Information Sciences, 8, Springer Verlag, Berlin (2^{nd} ed 1988).

LUTTREL S.P. (1994) : A bayesian analysis of self-organizing maps. Neural comput. 6.

OJA E. and S. KASKI (1999) : Kohonen maps. Elsevier

SCHROEDER A. (1976): Analyse d'un mélange de distribution de probabilité de même type. RSA. vol. 24, n^0 1.

THIRIA S., LECHEVALLIER Y., GASCUEL O. et CANU S., (1997) : Statistique et méthodes neuronales. Dunod