Quantifying the Ocean's CO2 budget with a CoHeL-IBR system

Juan M. Corchado[1], Jim Aiken[2], Emilio S. Corchado[3], Nathalie Lefevre[4], and Tim Smyth[2]

[1]Departamento Informática y Automática
Universidad de Salamanca
Plaza de la Merced s/n
37008, Salamanca, Spain
corchado@usal.es
[2]Centre for Air-Sea Interactions and fluxes
Plymouth Marine Laboratory, Prospect Place,
Plymouth, PL1 3 DH, U.K.
ja@mail.pml.ac.uk
[3]Departamento de Ingeniería Civil, Escuela Politécnica Superior.
Universidad de Burgos.
C/ Francisco de Vitoria s/n
09006, Burgos, Spain
escorchado@ubu.es
[4]School of Environmental Sciences
University of East Anglia
Norwich, NR4 7TJ, U.K.
n.lefevre@ueaS.ac.uk

Abstract. By improving accuracy in the quantification of the ocean's CO_2 budget, a more precise estimation can be made of the terrestrial fraction of global CO_2 budget and its subsequent effect on climate change. First steps have been taken towards this from an environmental and economic point of view, by using an instance based reasoning system, which incorporates a novel clustering and retrieval method - a Cooperative Maximum Likelihood Hebbian Learning model (CoHeL). This paper reviews the problems of measuring the ocean's CO_2 budget and presents the CoHeL model developed and outlines the IBR system developed to resolve the problem.

1 Introduction

This paper presents the results obtained with an instance based reasoning system (IBR) developed to estimate the ocean-atmosphere partial pressure of CO_2 (pCO_2) from information extracted from satellite pictures, wind direction and strength and other parameters such as water temperature, salinity and fluorescence. The final goal of our project is to construct a model that calculates the exchange rate and the global

budgets of CO_2, between the ocean and the atmosphere. An understanding of the natural sources and sinks of atmospheric carbon dioxide is necessary for predicting future atmospheric loading and its consequences for global climate. Present estimates of emissions and uptake do not balance, and although some have attributed the imbalance to a terrestrial sink, the magnitude of the oceanic sink remains undefined [7]. The rapid increase in atmospheric CO_2 resulting from atmospheric changes in the carbon cycle has stimulated a great deal of interest. Important decisions need to be made about future tolerable levels of atmospheric CO_2 content, as well as the lead and fossil fuel usage strategies that will permit us to achieve our environmental goals. The solution to these types of problems requires the use of dynamic systems, capable of incorporating new knowledge and facilitating the monitoring and estimation work carried out by oceanographers [14].

Case based reasoning (CBR) and IBR systems have been successfully used in several domains such as diagnosis, prediction, control and planning [13] [5] [18]. However, a major problem with these systems is the difficulty of case retrieval and case matching when the number of cases increases; large case bases are difficult to handle and require efficient indexing mechanisms and optimised retrieval algorithms. Moreover, there are very few standard techniques for automating their construction, since each problem may be represented by a different data set and requires a customised solution. Based on recent successful experiments with this technology [6] an instance based reasoning system has been developed for estimating the partial pressure of CO_2 in the ocean. The IBR system developed incorporates a novel Cooperative Maximum Likelihood Hebbian Learning model for the data clustering and retrieval and a radial-bases function neural network for instance adaptation and forecast, which is an extension and an improvement of the one presented in [6].

This paper reviews a method that can be used for the automation of IBR systems especially developed for estimating the partial pressure of CO_2 in an area of the Pacific ocean from Latitude 22,6ºS to 24ºS and Longitude 70ºW to 72ºW, which corresponds to a water mass situated off the Chile coasts of “Mejillones” and “Antofagasta”. The Cooperative Maximum Likelihood Hebbian Learning (CoHeL) method is a novel approach that features both selection, in which the aim is to visualize and extract information from complex, and highly dynamic data. The model proposed is a mixture of factor analysis and exploratory projection pursuit [8] based on a family of cost functions proposed by Fyfe and Corchado [9] which maximizes the likelihood of identifying a specific distribution in the data while minimizing the effect of outliers [9] [16]. It employs cooperative lateral connections derived from the Rectified Gaussian Distribution [3] [15] in order to enforce a more sparse representation in each weight vector. This method is used for the clustering of instances, and during the retrieval stage of the IBR cycle, the adaptation step is carried out using a radial basis function network while the revision stage is manually carried out by an oceanographer (since the specific aim of this project is to construct a tool for oceanographers). Finally, the system is updated continuously with data obtained from the afore mentioned satellites and sensors.

First, the present paper will describe the oceanographic problem that defines the framework of our research, then the CoHeL method, used to automate the retrieval stage of the IBR systems, will be described. A presentation will then be made of the

instance based reasoning model and finally, the results of the experiments will be described.

2 Ocean-atmosphere interaction

The oceans contain approximately 50 times more CO_2 in dissolved forms than the atmosphere, while the land biosphere including the biota and soil carbon contains about 3 times as much carbon (in CO_2 form) as the atmosphere [17]. The CO_2 concentration in the atmosphere is governed primarily by the exchange of CO_2 with these two dynamic reservoirs. Since the beginning of the industrial era, about 2000 billion tons of carbon have been released into the atmosphere as CO_2 from various industrial sources including fossil fuel combustion and cement production. At present, atmospheric CO_2 content is increasing at an annual rate of about 3 billion tons which corresponds to one half of the annual emission rate of approximately 6 billion tons from fossil fuel combustion. Whether the missing CO_2 is mainly absorbed by the oceans or by the land and their ecosystems have been debated extensively over the past decade. It is important, therefore, to fully understand the nature of the physical, chemical and biological processes which govern the oceanic sink/source conditions for atmospheric CO2 [11] [17].

New satellite sensors: ENVISAT, Aqua and other new Earth Observation satellites herald a new era in marine Earth Observation. Satellite-borne instruments provide high-precision, high-resolution data on atmosphere, ocean boundary layer properties and ocean biogeochemical variables, daily, globally, and in the long term. All these new sources of information have changed our approach to oceanography and the data generated needs to be fully exploited. Wind stress, wave breaking and the damping of turbulence and ripples by surface slicks, all affect the air-sea exchange of CO_2. These processes are closely linked to the "roughness" of the sea surface, which can be measured by satellite radars and microwave radiometers. Sea surface roughness consists of a hierarchy of smaller waves upon larger waves (photograph, left, and close-up, below). Different sensors give subtly different measurements of this roughness. Our final aim is to model both the open ocean and shelf seas, and it is believed that by assimilating Earth Observation (EO) data into artificial intelligence models these problems may be solved. EO data (both for assimilation and for validation) are vital for the successful development of reliable models that can describe the complex physical and biogeochemical interactions involved in marine carbon cycling. Satellite information is vital for the construction of oceanographic models, and in this case, to produce estimates of air-sea fluxes of CO_2 with much higher spatial and temporal resolution, using artificial intelligence models than can be achieved realistically by direct *in situ* sampling of upper ocean CO_2.

The systems have been tested in a number of cruises carried out off Chile during the austral summer of 2000, such as the one shown in Figure 1. The oceanographic cruises had several purposes including the calibration of new satellites and sensors, evaluation of the model proposed, etc. During the cruise, data was obtained *in situ* from temperature, chlorophyll, fluorescence and salinity sensors, and satellite images were also obtained. Partial pressure of CO_2 (pCO_2) was also calculated in real time.

This data was used to calibrate satellite sensors and to feed the IBR system, with the intention of developing a model that may allow, in the future, the calculation of pCO_2 values from satellite images rather than from *in situ* cruises.

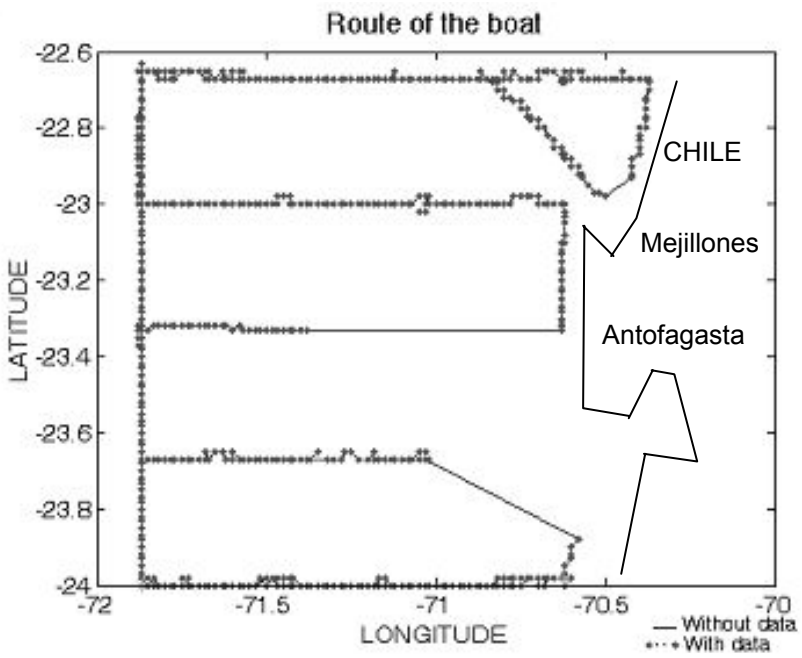

Fig. 1. Cruise track in the Pacific waters.

3 CoHeL Model

The Cooperative Maximum Likelihood Hebbian Learning (CoHel) method used during the retrieval stage of an IBR system is closely related to factor analysis and exploratory projection pursuit. It is a neural model based on the Negative Feedback artificial neural network, which has been extended by the combination of two different techniques. As mentioned before, this method is an extension and an improvement of the Maximum Likelihood Hebbian Learning (MLHL) method presented in [6]. This is a more robust method, based on a more solid mathematical formulation and that enforces a more sparse representation due to the use of cooperative lateral connections among the output neurons of the neural net, which also guaranties a faster convergence and clustering.

In this case, after selecting a cost function from a family of cost functions which identify different distributions - this method is called Maximum-Likelihood Hebbian learning [4][9] - cooperative lateral connections derived from the Rectified Gaussian Distribution [15] were added to the Maximum-Likelihood method which enforced a

greater sparsity in the weight vectors. To understand the proposed method is necessary to review the concepts presented in the following sections.

3.1 The Negative Feedback Neural Network

First, we shall present the Negative Feedback Network, which is the basis of the Maximum-Likelihood model. Feedback is said to exist in a system whenever the output of an element in the system partially influences the input applied to that particular element. It is used in this case to maintain the equilibrium on the weight vectors.

Consider an N-dimensional input vector, $\mathbf{x}$, and a M-dimensional output vector, $\mathbf{y}$, with W_{ij} being the weight linking input j to output i and let η be the learning rate. The initial situation is that there is no activation at all in the network. The input data is fed forward via weights from the input neurons (the $\mathbf{x}$ -values) to the output neurons (the $\mathbf{y}$ -values) where a linear summation is performed to give the activation of the output neuron. We can express this as:

$$y_i = \sum_{j=1}^{\mathbf{N}} W_{ij} x_j, \quad \forall i \tag{1}$$

The activation is fed back through the same weights and subtracted from the inputs (where the inhibition takes place):

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \quad \forall j \tag{2}$$

After that simple Hebbian learning is performed between input and outputs:

$$\Delta W_{ij} = \eta e_j y_i \tag{3}$$

The effect of the negative feedback is to stabilise the learning in the network. Because of this, it is not necessary to normalise or clip the weights to get convergence to a stable solution.

Note that this algorithm is clearly equivalent to Oja's Subspace Algorithm [12] since if we substitute Equation 2 in Equation 3 we get:

$$\Delta W_{ij} = \eta e_j y_i = \eta \left(x_j - \sum_k W_{kj} y_k \right) y_i \tag{4}$$

This network is capable of finding the principal components of the input data in a manner that is equivalent to Oja's Subspace algorithm [12], and so the weights will not find the actual Principal Components but a basis of the Subspace spanned by these components. Since the model is equivalent to Oja's Subspace algorithm, we might legitimately ask what we gain by using the negative feedback in such a way. Writing the algorithm like this gives us a model of the process which allows us to envisage different models which would otherwise be impossible [9].

Factor Analysis is a technique similar to PCA in that it attempts to explain the data set in terms of a smaller number of underlying factors. However Factor Analysis

begins with a specific model and then attempts to explain the data by finding parameters which best fit this model to the data. Charles and Fyfe [2] have linked a constrained version of the Negative Feedback network to Factor Analysis. The constraint put on the network was a rectification of either the weights or the outputs (or both). Thus if the weight update resulted in negative weights, those weights were set to zero; if the feed forward mechanism gives a negative output, this was set to zero. We will use the notation $[t]^+$ for this rectification: if t<0, t is set to 0; if t>0, t is unchanged.

3.2 ε-Insensitive Hebbian Learning

It has been shown that the nonlinear PCA rule

$$\Delta W_{ij} = \eta\left(x_j f(y_i) - f(y_i)\sum_k W_{kj} f(y_k) \right) \tag{5}$$

can be derived as an approximation to the best non-linear compression of the data. Thus we may start with a cost function

$$J(W) = 1^T E\left\{ \left(\mathbf{x} - Wf\left(W^T\mathbf{x}\right)\right)^2 \right\} \tag{6}$$

which we minimise to get the rule(5). Fyfe and MacDonald [10] used the residual in the linear version of (6) to define a cost function of the residual

$$J = f_1(\mathbf{e}) = f_1\left(\mathbf{x} - W\mathbf{y}\right) \tag{7}$$

where $f_1 = \|.\|^2$ is the (squared) Euclidean norm in the standard linear or nonlinear PCA rule. With this choice of $f_1(\ )$, the cost function is minimized with respect to any set of samples from the data set on the assumption that the residuals are chosen independently and identically distributed from a standard Gaussian distribution. We may show that the minimization of J is equivalent to minimizing the negative log probability of the residual, **e**, if **e** is Gaussian. Let:

$$p(\mathbf{e}) = \frac{1}{Z}\exp(-\mathbf{e}^2) \tag{8}$$

The factor Z normalizes the integral of $p(\mathbf{y})$ to unity.

Then we can denote a general cost function associated with this network as

$$J = -\log p(\mathbf{e}) = (\mathbf{e})^2 + K \tag{9}$$

where K is a constant. Therefore performing gradient descent on J we have

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx \mathbf{y}(2\mathbf{e})^T \tag{10}$$

where a less important term has been has been discarded. In general, the minimisation of such a cost function may be thought to make the probability of the residuals more dependent on the probability density function (pdf) of the residuals [16]. Thus if the probability density function of the residuals is known, this

knowledge could be used to determine the optimal cost function. Fyfe and MacDonald [10] investigated this with the (one dimensional) function:

$$p(\mathbf{e}) = \frac{1}{2+\varepsilon}\exp\left(-|\mathbf{e}|_{\varepsilon}\right) \tag{11}$$

where

$$|\mathbf{e}|_{\varepsilon} = \begin{cases} 0\_\forall|\mathbf{e}| < \varepsilon \\ |\mathbf{e}| - \varepsilon\_otherwise \end{cases} \tag{12}$$

with ε being a small scalar ≥ 0. Fyfe and Corchado [9] described this in terms of noise in the data set. However we feel that it is more appropriate to state that, with this model of the pdf of the residual, the optimal $f_1(\ )$ function is the ε-insensitive cost function:

$$f_1(\mathbf{e}) = |\mathbf{e}|_{\varepsilon} \tag{13}$$

In the case of the Negative Feedback Network, the learning rule is

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial f_1(\mathbf{e})}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \tag{14}$$

which gives:

$$\Delta W_{ij} = \begin{cases} 0\_if|e_j| < \varepsilon \\ \eta y_i(sign(e_j)) \text{otherwise} \end{cases} \tag{15}$$

The difference with the common Hebb learning rule is that the sign of the residual is used instead of the value of the residual. Because this learning rule is insensitive to the magnitude of the input vectors x, the rule is less sensitive to outliers than the usual rule based on mean squared error. This change from viewing the difference after feedback as simply a residual rather than an error will permit us later to consider a family of cost functions each member of which is optimal for a particular probability density function associated with the residual.

3.3 A Family of Learning Rules

Now the ε-insensitive learning rule is clearly only one of a possible family of learning rules which are suggested by the family of exponential distributions. Let the residual after feedback have probability density function

$$p(\mathbf{e}) = \frac{1}{Z}\exp(-|\mathbf{e}|^p) \tag{16}$$

Then we can denote a general cost function associated with this network as

$$J = -\log p(\mathbf{e}) = |\mathbf{e}|^p + K \tag{17}$$

where K is a constant. Therefore performing gradient descent on J we have

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx y(p \mid \mathbf{e} \mid^{p-1} sign(\mathbf{e}))^T \tag{18}$$

where T denotes the transpose of a vector. We would expect that for leptokurtotic residuals (more kurtotic than a Gaussian distribution), values of p<2 would be appropriate, while for platykurtotic residuals (less kurtotic than a Gaussian), values of p>2 would be appropriate. It is a common belief in the ICA community that it is less important to get exactly the correct distribution when searching for a specific source than it is to use a model with an approximately correct distribution i.e. all supergaussian signals can be retrieved using a generic leptokurtotic distribution and all subgaussian signals can be retrieved using a generic platykurtotic distribution. The experiments [3] tend to support this belief to some extent but we often find accuracy and speed of convergence are improved when we are accurate in our choice of p.

Therefore the network operation is:

Feedforward: $$y_i = \sum_{j=1}^{N} W_{ij} x_j, \ \forall_i \tag{19}$$

Feedback: $$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i \tag{20}$$

Weight change: $$\Delta W_{ij} = \eta . y_i . sign\left(e_j\right) \mid e_j \mid^{p-1} \tag{21}$$

Corchado and Fyfe [3] described their rule as performing a type of PCA, but this is not strictly true since only the original (Oja) ordinary Hebbian rule actually performs PCA. It might be more appropriate to link this family of learning rules to Principal Factor Analysis since this method makes an assumption about the noise in a data set and then removes the assumed noise from the covariance structure of the data before performing a PCA. We are doing something similar here in that we are basing our PCA-type rule on the assumed distribution of the residual. By maximising the likelihood of the residual with respect to the actual distribution, we are matching the learning rule to the pdf of the residual.

This method has been linked to the standard statistical method of Exploratory Projection Pursuit (EPP) [8]: EPP also gives a linear projection of a data set but chooses to project the data onto a set of basis vectors which best reveal the interesting structure in the data; interestingness is usually defined in terms of how far the distribution is from the Gaussian distribution.

3.4 Rectified Gaussian Distribution

The Rectified Gaussian Distribution [15] is a modification of the standard Gaussian distribution in which the variables are constrained to be non-negative, enabling the use of non-convex energy functions. The multivariate normal distribution can be defined in terms of an energy or cost function in that, if realised samples are taken far from the distribution's mean, they will be deemed to have high energy and this will be equated to low probability. More formally, we may define the standard Gaussian distribution by:

$$p(\mathbf{y}) = Z^{-1} e^{-\beta E(\mathbf{y})}, \tag{22}$$

$$E(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T \mathbf{A}\mathbf{y} - \mathbf{b}^T \mathbf{y} \tag{23}$$

The quadratic energy function $E(\mathbf{y})$ is defined by the vector **B** and the symmetric matrix **A**. The parameter $\beta = 1/T$ is an inverse temperature. Lowering the temperature concentrates the distribution at the minimum of the energy function.

One advantage of this formalisation is that it allows us to visualise regions of high or low probability in terms of energy and hence to view movement to low energy regions as movement to regions of high probability.

The quadratic energy function $E(\mathbf{y})$ can have different types of curvature depending on the matrix **A**. Consider the situation in which the distribution of the firing of the outputs of our neural network follows a Rectified Gaussian Distribution. Then it is possible to identify values of **A** which give increasingly sparse firings and in the extreme, a single neuron will respond to the whole data set. Two examples of the Rectified Gaussian Distribution are the competitive and cooperative distributions. The modes of the competitive distribution are well-separated by regions of low probability. The modes of the cooperative distribution are closely spaced along a non-linear continuous manifold. Our experiments focus on a network based on the use of the cooperative distribution.

Neither distribution can be accurately approximated by a single standard Gaussian. Using the Rectified Gaussian, it is possible to represent both discrete and continuous variability in a way that a standard Gaussian cannot.

Not all energy functions can be used in the Rectified Gaussian Distribution. The sorts of energy function that can be used are only those where the matrix A has the property:

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} > \mathbf{0} \text{ for all } \mathbf{y} : y_i > 0, i = 1 \ldots N \tag{24}$$

where N is the dimensionality of y. This condition is called co-positivity. This property blocks the directions in which the energy diverges to negative infinity. The cooperative distribution in the case of N variables is defined by:

$$A_{ij} = \delta_{ij} + \frac{1}{N} - \frac{4}{N}\cos\left(\frac{2\pi}{N}(i-j)\right) \tag{25}$$

$$b_i = 1 \tag{26}$$

where δ_{ij} is the Kronecker delta and *i* and *j* represent the identifiers of output neuron. To speed learning up, the matrix **A** can be simplified [3] to:

$$A_{ij} = \left(\delta_{ij} - \cos\left(2\pi(i-j)/N\right)\right) \tag{27}$$

The matrix **A** is used to modify the response to the data based on the relation between the distances between the outputs. The outputs are thought of as located on a ring ("wraparound"). Note that the modes of the Rectified Gaussian are the minima of the energy function, subject to non-negativity constraints. The modes of the

distribution characterize much of its behaviour at low temperature. Finding the modes of a Rectified Gaussian is a problem in quadratic programming. However we will use what is probably the simplest algorithm, the projected gradient method, consisting of a gradient step followed by a rectification:

$$y_i(t+1) = \left[y_i(t) + \tau(b - \mathrm{Ay})\right]^+ \tag{28}$$

where the rectification $[\ ]^+$ is necessary to ensure that the y-values keep to the positive quadrant. If the step size τ is chosen correctly, this algorithm can probably be shown to converge to a stationary point of the energy function [1].

In practice, this stationary point is generally a local minimum. The mode of the distribution can be approached by gradient descent on the derivative of the energy function with respect to **y**. This is:

$$\Delta\mathbf{y} \propto -\frac{\partial E}{\partial \mathbf{y}} = -(\mathrm{A}\mathbf{y} - \mathbf{b}) = \mathbf{b} - A\mathbf{y}\ ; \tag{29}$$

which is used as in Equation 28.

Now the rectification in Equation 28 is identical to the rectification which Corchado and Fyfe [3] used in the Maximum-Likelihood Network. Thus we will use this movement towards the mode in the Factor Analysis version of the Maximum-Likelihood Network before training the weights as previously. The net result will be shown to be a network which can find the independent factors of a data set but do so in a way which captures some type of global ordering in the data set.

We use the standard Maximum-Likelihood Network but now with a lateral connection (which acts after the feed forward but before the feedback). Thus we have

Feedforward: $$y_i = \sum_{j=1}^{\mathbf{N}} W_{ij} x_j, \ \forall i \tag{30}$$

Lateral Activation Passing: $$y_i(t+1) = \left[y_i(t) + \tau(b - Ay)\right]^+ \tag{31}$$

Feedback: $$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \ \forall j \tag{32}$$

Weight change: $$\Delta W_{ij} = \eta . y_i . sign(e_j) |\ e_j\ |^{p-1} \tag{33}$$

Where the parameter τ represents the strength of the lateral connections.

4 A CoHeL-IBR system for calculating the exchange rate of CO_2

An IBR system has been constructed for obtaining the value of the exchange rate or surface partial pressure of CO_2 (pCO_2) in oceanographic waters from biological parameters and satellite information. The IBR system uses the Cooperative Maximum Likelihood Hebbian Learning Model for clustering the Instance-base and for the retrieval of the instances most similar to the "problem instance", due to its topology preserving properties. The selected instances are used during the reuse stage to train a radial function neural network [5] [9], that provides the value of the pCO_2 for a given

point and the result is evaluated by an oceanographer. The learning (retain stage) is carried out by updating the instance base, updating the weights of the radial basis function network and by re-calling the Cooperative Maximum Likelihood Hebbian Learning Model for the clustering of the data.

Table 1. Instance attributes.

Instance Field	Measurement
JD	Serial day of the year
LAT	Latitude
LONG	Longitude
SST	Temperature
S	Salinity
WS	Wind strength
WD	Wind direction
Fluo_calibrated	fluorescence calibrated with chlorophyll
SW pCO2	surface partial pressure of CO2

Applying equations 30 to 33 to the instance-base, the algorithm automatically groups the instances into clusters, grouping together those of similar structure. This technique is a classification and visualisation tool for high dimensional data on a low dimensional display. One of the advantages of this technique is that it is an unsupervised method so we do not need to have any information about the data beforehand. When a new instance is presented to the IBR system, it is identified as belonging to a particular type by applying equation 30 to it. Each stored instance contains information relating to a specific situation.

Table 1 presents the values used to define the problem. Where JD, LAT, LONG, SST, S, WS, WD, Fluo_calibrated and SW represent the problem description and pCO2 is the value that the IBR system has to identify from the problem descriptor. These values for a given point can be obtained from cruises using sensors or from satellite images. Initially the system was tested *in situ* during the cruise carried out in Pacific waters. The instance-base of the system was fed with 85% of the instances recorded during the cruise (over 85.000 instances). The other 15%, homogeneously spread along the cruise track, was left in order to test the system after the cruse was completed.

The results obtained were very accurate, with an average error of 7,4%, which is less than the error provided by the other techniques we used to evaluate the IBR system. Table 2 presents the average error obtained with the CoHeL-IBR system, with the MLHL-IBR system [6], a Radial-basis Function Neural Network, a Multi-layer Perceptron Neural Network, a Growing Cell Structures Neural Network and a K-nearest neighbour algorithm.

Starting from the error series generated by the different models, the Kruskall-Wallis test has been carried out. Since the P-value is less than 0,01, there is a statistically significant difference between the models at the 99,0% confidence level. Table 3 shows a multiple comparison procedure (Mann-Withney test) used to determine which models are significantly different from the others. The asterisk indicates that these pairs show statistically significant differences at the 99.0%

confidence level. Table 3 shows that the IBR system presents statistically significant differences from the other models. The proposed model generates the best results of all the tested techniques. Figure 3 presents the error obtained in 40 cases in with the system was tested. These cases have been randomly obtained from the testing data set (15% of the whole data set), the other 85% of the data set was used to create the model.

Table 2. Average error obtained with the IBR system and other methods.

Method	Average Error
CoHeL-IBR system	7,4%
MLHL-IBR system	8,2%
Radial-basis Function Network	9,8%
Multi-layer Perceptron	10,1%
Growing Cell Structures	16,2%
K-nearest neighbour	13,6%

Table 3: Mann-Withney test results.

	CoHeL-IBR	MLHL-IBR	RBF	MLP	GCS	KNN
CoHeL-IBR						
MLHL-IBR	*					
RBF	*	*				
MLP	*	*	=			
GCS	*	*	*	*		
KNN	*	*	*	=	*	

The final goal of the project is to calculate the value of pCO2 from satellite images, such as the ones shown in Figure 2. Most of the values of the parameters presented in Table 1 can be directly obtained from such photographs and others may be extracted with some well-known calculation. In this case, the CoHeL-IBR system was tested with data extracted from satellite images of the area in which the cruise took place, such as the ones presented in Figure 2. Problem instances (vectors with the values of: JD, LAT, LONG, SST, S, WS, WD, Fluo_calibrated and SW) were constructed, along the cruise track from such images and were fed into the CoHeL-IBR system, in order for it to obtain the value of the pCO_2. In this case the average error of the CoHeL-IBR system was slightly higher, but still very accurate compared with the results obtained with the other techniques. Oceanographers have also consider these results to be highly significant. The second column of Table 4 shows these results. Then problem instances were obtained from the same photographs, but from points outside the cruise tracks, and similar results were obtained, as shown in the third column of Table 4.

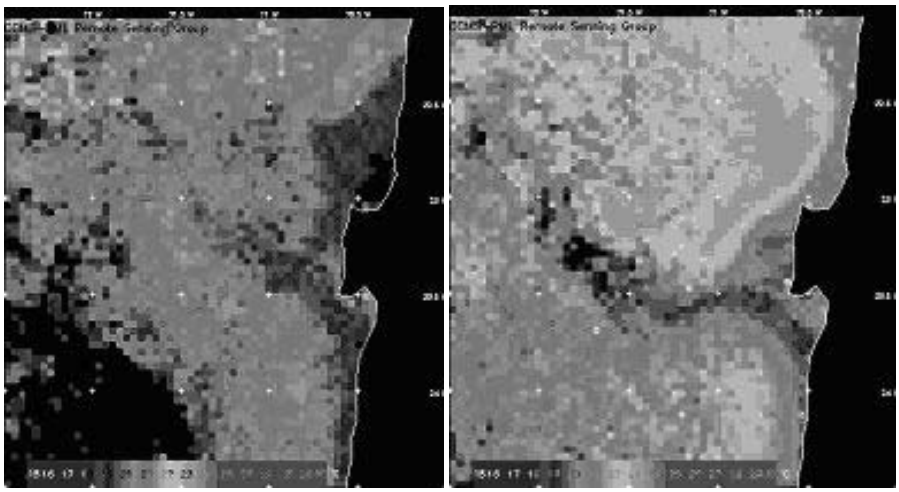

Fig. 2. Processed NOAA Satellite images, showing temperature values, obtained at one week intervals.

Table 4. Average error obtained with the IBR system and other methods on Satellite data.

Method	Average error (Track data)	Average error (Out side track data)
CoHeL-IBR	9,7%	10,3%
MLHL-IBR	12,3%	13,9%
RBF	13,1%	14,5%
MLP	15,2%	14,7%
GCS	18,9%	18,8%
KNN	17,2%	18,1%

5 Conclusions and Future Work

The CoHeL-IBR system presented is able to produce a forecast with an acceptable degree of accuracy. The final constructed tool constitutes the first system developed for calculating the pCO_2 *in situ* and from satellite images. The IBR system incorporates a novel clustering technique capable of indexing huge instance-bases in an unsupervised way and of successfully retrieving instances with a similar structure, which is vital for constructing a model with a radial basis function neural network.

The Cooperative Maximum Likelihood Hebbian Learning Model has also performed better that other algorithms, due to its fast convergence and clustering abilities. It enforces a more sparse representation due to the use of cooperative lateral connections among the output neurons of the neural net. With this technique, the retrieval of the best matching instance is a very simple operation using the proposed method and presents no major computational obstacles. The proposed method is both

advantageous in the creation of and retrieval from instance bases but is also important in its own right in the unsupervised investigation of interesting structure in high dimensional data sets. The results obtained in both experiments are very encouraging and the model presents great potential. The experiments carried out have allowed us to determine the efficiency of the model when the data used to create the instance-base and the problem instances is reliable. It has also been shown the potential of the model to automate the resolution of the problem with the help of satellite photographs. In this case, the error may be due to calibration imbalances, lack of definition of the photographs, presence of clouds, errors in the wind measures, etc. These are some of the problems that have to be solved in the framework of this project. Table 4 also shows the generalization capabilities of the proposed model, since it is even able to generate reasonable results in an extended area, when the instance-base has only been constructed with data from one part of the area. More experiments need to be carried out for the model validation and techniques to facilitate the revision of the solution have to be obtained. The uncertainty and the dynamism of oceanographic systems have to be taken into consideration and techniques for monitoring such factors need to be incorporated into the system. The proposed model is a first step towards the resolution of this complex problem, which still requires a great deal more work and research.

References

1. Bertsekas D.P. (1995) Nonlinear Programming. Athena Scientific, Belmont, MA, 1995.
2. Charles D. and Fyfe C. (1998) Modeling Multiple Cause Structure Using Rectification Constraints. Network: Computation in Neural Systems, 9:167-182.
3. Corchado E. and Fyfe C. (2003) Orientation Selection Using Maximum Likelihood Hebbian Learning, International Journal of Knowledge-Based Intelligent Engineering Systems Volume 7 Number 2, April 2003.
4. Corchado E., MacDonald D. and Fyfe C. (2004) Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit, Data mining and Knowledge Discovery, Kluwer Academic Publishing, (In press).
5. Corchado J. M. and Lees B. (2001) A hybrid case-based model for forecasting. Applied Artificial Intelligence: An International Journal Vol 15, no. 2, pp 105-127.
6. Corchado J. M., Corchado E. S., Aiken J., Fyfe C., Fdez-Riverola F. and Glez-Bedia M. (2003) Maximum Likelihood Hebbian Learning Based Retrieval Method for CBR Systems. 5th International Conference on Case-Based Reasoning, Trondheim, Norway, June 23 to 26, 2003. Springer-Verlag
7. Finnegan W. G., Pitter R. L. and Young L. G. (1991) Atmospheric Environment. A25, 2531-2534.
8. Freedman J. and Tukey J. (1974) A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, (23): 881-890, 1974.
9. Fyfe C. and Corchado E. S. (2002) Maximum Likelihood Hebbian Rules. European Symposium on Artificial Neural Networks. 2002.

10. Fyfe C. and MacDonald D. (2002), ε-Insensitive Hebbian learning, Neuro Computing, 2002
11. Lefevre N., Aiken J., Rutllant J., Daneri G., Lavender S. and Smyth T. (2002) Observations of pCO2 in the coastal upwelling off Chile: Sapatial and temporal extrapolation using satellite data. Journal of Geophysical research. Vol. 107, no. 0
12. Oja E. (1989) Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, 1:61-68.
13. Pal S. K., Dillon T. S. and Yeung D. S. (2000) Soft Computing in Case-based Reasoning. (eds.). Springer Verlag, London, U.K.
14. Sarmiento J. L. and Dender M. (1994) Carbon biogeochemistry and climate change. Photosynthesis Research, Vol. 39, 209-234.
15. Seung H.S., Socci N.D. and Lee D. (1998) The Rectified Gaussian Distribution, Advances in Neural Information Processing Systems, 10.
16. Smola A.J. and Scholkopf B. (1998) A Tutorial on Support Vector Regression. Technical Report NC2-TR-1998-030, NeuroCOLT2 Technical Report Series.
17. Takahashi T., Olafsson J., Goddard J. G., Chipman D. W. and Sutherland S. C. (1993) Seasonal Variation of CO_2 and nutrients in the High-latitude surface oceans: a comparative study. Global biochemical Cycles. Vol. 7, no. 4. pp 843-878.
18. Watson I. and Marir F. (1994) Case-Based Reasoning: A Review. Cambridge University Press, 1994. The knowledge Engineering Review. Vol. 9. Nº3.