The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2022-38

Workshop on Theoretical Ecology and Global Change

2 - 18 March 2009

Artificial Neural networks and remote sensing in the analysis of the highly variable pampean shallow lakes

CANZIANI Graciela Ana
Universidad Nacional del Centro de La Provincia de Buenos Aires
Facultad de Ciencias Exactas Departamento de Matematica
Campus Paraje Arroyo Seco, 7000 Tandil
ARGENTINA

MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **5**, Number **4**, October **2008**

doi:10.3934/mbe.2008.5.691

pp. **691–711**

ARTIFICIAL NEURAL NETWORKS AND REMOTE SENSING IN THE ANALYSIS OF THE HIGHLY VARIABLE PAMPEAN SHALLOW LAKES

GRACIELA CANZIANI, ROSANA FERRATI, CLAUDIA MARINELLI, FEDERICO DUKATZ
Multidisciplinary Institute on Ecosystems and Sustainable Development
Universidad Nacional del Centro de la Provincia de Buenos Aires
Pinto 399, 7000 Tandil, Argentina

Abstract. Suspended organic and inorganic particles, resulting from the interactions among biological, physical, and chemical variables, modify the optical properties of water bodies and condition the trophic chain. The analysis of their optic properties through the spectral signatures obtained from satellite images allows us to infer the trophic state of the shallow lakes and generate a real time tool for studying the dynamics of shallow lakes. Field data (chlorophyll-a, total solids, and Secchi disk depth) allow us to define levels of turbidity and to characterize the shallow lakes under study. Using bands 2 and 4 of LandSat 5 TM and LandSat 7 ETM+ images and constructing adequate artificial neural network models (ANN), a classification of shallow lakes according to their turbidity is obtained. ANN models are also used to determine chlorophyll-a and total suspended solids concentrations from satellite image data. The results are statistically significant. The integration of field and remote sensors data makes it possible to retrieve information on shallow lake systems at broad spatial and temporal scales. This is necessary to understanding the mechanisms that affect the trophic structure of these ecosystems.

1. **Introduction.** More than 10,000 permanent or temporary shallow lakes exist in the province of Buenos Aires, Argentina [5]. They form a system of wetlands that covers more that one million hectares, distributed in both permanent and temporary water bodies called "lagunas". Their surfaces vary in the range of 10 to 40,000 ha, following the alternation of humid and dry periods characteristic of the Pampean climate.

The numerous lentic water bodies characteristic of the Pampean region are generated by very low slopes and natural obstruction to drainage, together with the processes of wind deflation that occurred during the quaternary period, and the present humid climate [8].

The Pampean "lagunas" are defined as shallow flatland lakes, polymictic, eutrophic and having highly variable water permanence time and a high biomass in their biotic communities [24]. These wetlands do not exhibit thermal stratification. Normally they are subject to periods of floods and droughts that change their volume. Both water and sediments are disturbed by winds. This generates a constant resuspension of sediments that releases nutrients and reduces oxygen in the water column. These shallow lakes have highly variable salinity and turnover rates, are

2000 *Mathematics Subject Classification.* Primary: 68T05, Secondary: 92D40.

Key words and phrases. eutrophic shallow lakes, remote sensing, artificial neural network.

This work was financed by Argentina's National Agency for the Promotion of Science and Tecnology, ANPCyT, BID 1201/OC-AR, through project PICTO 13-502.

naturally eutrophic, and suffer from environmental stress which increases even more their nutrient contents.

These conditions generate wide ranges of values of suspended sediments, phytoplankton, and dissolved organic matter, resulting in different turbidity states that modify the optic characteristics of the water. Most important, this also conditions both the trophic type and the biogeochemical fluxes in the shallow lakes [31].

Within this generality, the shallow lakes exhibit stages of clear waters and of turbid waters depending on the phytoplanktonic biomass, the presence of macrophytes, and the concentration of suspended inorganic solids. Shallow lakes in the clear water stage are more transparent and hence light penetration is deeper. This allows the development of submerged macrophytes. Various mechanisms related to the presence of macrophytes, such as shading, reduction in resuspension by wind effects, reduction in nutrient availability, and abundance of zooplankton which enhances grazing, usually combine to produce a reduction in chlorophyll [31]. In these clear shallow lakes there is a dominance of fish related to the community of submerged macrophytes and coexisting organisms, which facilitate the fishes' laying in wait and other activities as visual predators.

Shallow lakes in the turbid water stage are characterized by a biomass increase in all the pelagic communities due to a higher availability of nutrients that are taken by phytoplankton to the detriment of macrophytes. Moreover, different trophic states can be observed depending on the relative concentrations of suspended organic and inorganic matter in the water column. In the so called *green* shallow lakes, water turbidity is mainly caused by the presence of abundant phytoplanktonic biomass, while the *brow-* shallow lakes are characterized by a higher concentration of inorganic suspended matter arising from the sediment as an effect of the wind or from the surrounding landscape due to land use in the basin.

The environmental heterogeneity has lead to a partial and non-extrapolatable knowledge of the Pampean shallow lakes. The complexity of these ecosystems in general, and of their trophic structure in particular, makes very difficult to generate one general model for the functioning of the Pampean shallow lakes. Given the high number of shallow lakes and the extension of their spatial distribution, the use of remote sensors and the processing of satellite images become adequate complements to an integrated study of these water bodies. Developing a tool for the remote observation of such an ecosystem is necessary for establishing reasonable scenarios for management and decision-making.

The optic properties of terrestrial objects and, in particular, water, are properly perceived by the TM and ETM+ sensors located in LandSat 5 and LandSat 7 satellites respectively. Moreover, the spatial resolution of LandSat imagery is adequate for capturing the reduced size of the majority of Pampean shallow lakes. Obviously, it is essential that the data obtained by remote sensors be calibrated with field data in order to obtain a supervised classification of the water bodies.

The sensors are capable of capturing information in the form of spectral signatures which can be processed and become input for models which integrate these data and field data. Three components –suspended sediment, phytoplankton biomass (chlorophyll-a concentration) and dissolved organic carbon– are also the major factors that control the spectral signatures of water bodies [11]. This allows a more thorough analysis of the dynamics of these shallow lakes and helps to adequately identify, on the basis of objective and reproducible information, the successive states that appear seasonally or cyclically.

There is abundant literature on the use of remote sensors combined with artificial neural networks for the study of chlorophyll and sediments in oceanic and coastal waters. Estimates of chlorophyll and sediments in inland lakes using remote sensing have been developed mostly through the use of techniques based on simple or multiple linear regression models. Liu and collaborators [16] review the performance of different sensors and models for quantifying water quality parameters and analyze the use of linear and nonlinear models to determine the relationship between in situ samples and their corresponding remotely sensed data.

Svab and collaborators [33] analyze the best band combinations from LandSat TM and ETM+ for quantifying chlorophyll-a and suspended sediment concentrations. They find appropriate correlation coefficients between band 3 and suspended solids and between band 4 and chlorophyll-a for values bellow 200 mg/m^3. Their results confirm that suspended sediment rich water hides the spectral characteristics of chlorophyll-a when at relatively low concentration (typical environmental concentrations are bellow 100 mg/m^3 Chl a), especially for coarse spectral resolution platforms such as Landsat. At higher chlorophyll-a concentration, chlorophyll spectral characteristics begin to overcome the inorganic suspended sediment reflectance characteristics. [33].

Simple linear regression models ([17] [1]), multiple linear regression models ([32] [25] [2]), and nonlinear regression models ([20] [10]) have been applied to chlorophyll-a and suspended sediment quantifications. Liu [16] states that linear regression models are invalid for inland and shallow coastal waters that have a higher concentration level of water constituents or a higher degree of optical complexity than clear oceanic waters. Also, the relationship between field samples and their corresponding remotely sensed data can be linear or non-linear, but are nearly always site-specific [16].

For this reason, nonparametric statistical techniques such as neural network analysis have been introduced to model the transfer function between chlorophyll and sediment concentrations and satellite-received radiance with a much higher accuracy than multiple regression analysis [14].

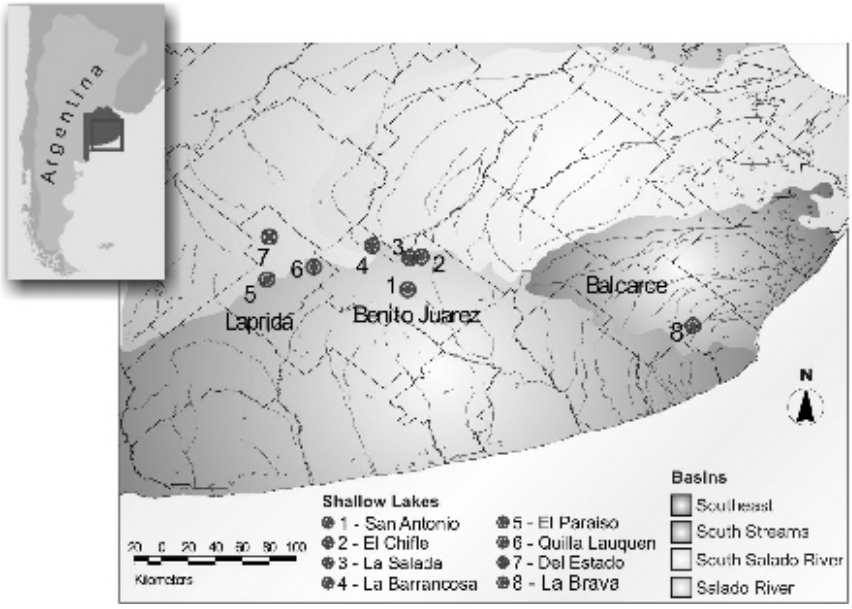
The general purpose of our research is to understand the structure and the functioning of the Pampean shallow lakes through the development of observation and monitoring tools. The spatial and temporal analysis will be based on the integration of data collected from the field and data retrieved from remote sensors, as well as the construction of adequate mathematical models.

Considering the possibly high number of samples that will be handled in an in-depth study of a given region, it is necessary to develop an automatic spatio-temporal analysis system with the capability of supervision. The tool presented here is part of such a system which also includes other modules for studying frequencies of appearance of temporary shallow lakes, recurrence of flooded areas, changes in individual lake area, spatial behavior of permanent water bodies [6], and spectral unmixing of components [7], among other tools.

The purpose of this paper is to determine if Landsat TM and ETM+ can be used to estimate trophic state parameters for shallow lakes in the Pampean region. The objectives are to develop and apply ANN models: (a) to the classification of Pampean shallow lakes following their turbidity, and (b) to relate Landsat TM and ETM+ reflectance to the trophic state parameters by quantifying chlorophyll-a and total solids concentrations.

2. Materials and methods.

2.1. **Location.** Eight shallow lakes located in different basins of the Province of Buenos Aires, Argentina, were selected for the development, calibration, and validation of the methodology here proposed. The *"lagunas"* or shallow lakes selected for this study are located in the districts of Juarez (San Antonio, La Salada, El Chifle and La Barrancosa), Balcarce (La Brava) and Laprida (El Paraiso, Quilla Lauquen, and Del Estado), on the water divide of the large hydrographic basins of the Province of Buenos Aires (Salado river, South streams, South-East streams, and South channels, [26]). (See Map 1)



MAP 1: Location of the study shallow lakes.

2.2. **Field data.** The field sampling was carried on between December 2004 and August 2006. The number of trips to each shallow lake were: eight for La Brava; twelve for Del Estado, Quilla Lauquen, and El Paraiso; sixteen for La Salada, El Chifle, and San Antonio; and seventeen for La Barrancosa. Field measurements performed in each of the shallow lakes included Secchi disk depth, and laboratory analyses provided chlorophyll-a and total solids concentrations, among other variables. Table 1 shows minimum, maximum, and mean values of Secchi disk depth, chlorophyll-a and total solids measured at each shallow lake.

The values of chlorophyll-a and total solids (TS) in the samples and the Secchi disk depth were used to perform a classification of the shallow lakes from field data following their turbidity. An initial classification was done by defining the quantile probability distribution [27]. Four categories were later defined from the quantile probability distribution and the expertise acquired in the field:

Class 1, formed by clear water shallow lakes
Class 2, formed by shallow lakes presenting intermediate turbidity values
Class 3, formed by shallow lakes exhibiting high values of chlorophyll

Class 4, formed by shallow lakes exhibiting the highest values of chlorophyll and total solids.

These categories were later used for training the artificial neural network (ANN). Table 2 shows mean, minimum, maximum, median, first and third quantile for each variable.

TABLE 1: Minimum, maximum, and mean values of chlorophyll-a and total solids measured at each shallow lake.

Shallow Lake	Secchi Disk Depth [cm] min-max (mean)	Chlorophyll *a* [mg/m^3] min-max (mean-median)	Total Solids [mg/l] min-max (mean-median)
La Brava	37–37 (37)	45.3-45.3 (45.3-45.3)	507–507 (507-507)
Quilla Lauquen	10–22 (16)	76.3-122.1 (99.2-99.2)	806–812 (809.0-809)
San Antonio	18–36 (26.2)	50.4-179.8 (101.4-94.0)	1296–1418 (1376.6-1396)
La Barrancosa	11.5–27 (19.5)	71.3-155.4 (110.5-95.7)	1327–1659 (1559.1-1624)
La Salada	13–23 (17.3)	112.0-260.2 (159.2-132.2)	1184–1284 (1241.3-1248.5)
El Chifle	7–23 (15.8)	120.9-249.9 (200.6-215.8)	1736–1953 (1846.8-1849)
El Paraiso	11–17 (14)	303.5-384.5 (353.9-369.0)	1498–1830 (1684.0-1700)
Del Estado	9–15.5 (9.5)	325.0-671.2 (559.5-600.8)	1788–2881 (2438.1-2522)

TABLE 2: Mean, minimum, maximum, median, first and third quantil values for chlorophyll-a, total solids (TS) and Secchi disk depth (SDD) defined for each class.

Class	Variable	Mean	Min	Max	Median	Q1	Q3
1	Chl *a*	60.80	45.30	76.30	60.80	45.30	76.30
	TS	659.50	507.00	812.00	659.50	507.00	812.00
	SDD	25.50	14.00	37.00	25.50	14.00	37.00
2	Chl *a*	130.03	50.40	260.20	117.50	93.20	155.40
	TS	1435.63	806.00	1953.00	1397.00	1284.00	1629.00
	SDD	20.63	11.50	36.00	19.00	17.00	23.00
3	Chl *a*	305.73	120.90	384.50	340.40	249.90	371.90
	TS	1728.29	1498.00	1942.00	1706.00	1686.00	1830.00
	SDD	13.71	10.00	17.00	15.00	11.00	16.00
4	Chl *a*	559.46	325.00	671.20	600.80	533.30	667.00
	TS	2438.20	1788.00	2881.00	2522.00	2347.00	2653.00
	SDD	9.00	7.00	10.00	9.00	9.00	10.00

2.3. **Remote sensors data.** The National Commission for Space Activities (Comision Nacional de Actividades Espaciales, CONAE) provided sixteen satellite images (224/86, 225/86, and 226/86), ten from LandSat 5 TM and six from LandSat 7 ETM+. The delay between the dates of the field trips and the satellite images was kept below six days [22]. Dates of satellite images and field samplings can be seen in Table 3.

One main concern when using remote sensing in shallow lakes is bottom interference. To make sure that the bottom did not influence the results, the shallow lakes were selected so that their depth was at least twice their Secchi disk depth [22], and hence reflectance from vegetation or the lake bottom would not affect the spectral signature (Table 4).

TABLE 3: Detail of the type and dates of images and dates of field sampling.

Images Path/Row Satellite Date	**Shallow Lake (area)**	**Field Sampling Date**
226/86 LandSat 5 TM 2004/12/14	La Barrancosa (181 ha)	2004/12/15
226/86 LandSat 5 TM 2005/03/04	La Barrancosa (172 ha) Del Estado (170 ha) El Paraíso (102 ha)	2005/02/28
225/86 LandSat 5 TM 2005/03/13	La Barrancosa (166 ha) San Antonio (139 ha)	2005/03/16
225/86 LandSat 5 TM 2005/04/22	La Barrancosa (172 ha) El Chifle (134 ha) La Salada (158 ha) San Antonio (182 ha)	2005/04/28
226/86 LandSat 5 TM 2005/05/07	Del Estado (162 ha) El Paraíso (91 ha) Quilla Lauquen (108 ha)	2005/05/05
226/86 LandSat 5 TM 2005/06/08	Quilla Lauquen (101 ha) Del Estado (147 ha) El Paraíso (96 ha)	2005/06/06
225/86 LandSat 5 TM 2005/07/11	La Barrancosa (164 ha) El Chifle (131 ha) La Salada (152 ha) San Antonio (174 ha) Quilla Lauquen (98 ha)	2005/07/07
226/86 LandSat 7 ETM 2005/07/18	El Paraíso (107 ha)	2005/07/13
226/86 LandSat 5 TM 2005/09/12	Del Estado (139 ha) El Paraíso (97 ha) Quilla Lauquen (104 ha) La Barrancosa (167 ha)	2005/09/07
225/86 LandSat 7 ETM 2005/09/13	San Antonio (175 ha) La Salada (188 ha) El Chifle (133 ha)	2005/09/07
225/86 LandSat 7 ETM 2005/10/15	San Antonio (171 ha) La Salada (151 ha) El Chifle (133 ha)	2005/10/14
226/86 LandSat 5 TM 2005/10/14	La Barrancosa (167 ha) Quilla Lauquen (102 ha) El Paraíso (96 ha) Del Estado (131 ha)	2005/10/14
225/86 LandSat 7 ETM 2005/12/02	La Barrancosa (158 ha) Quilla Lauquen (94 ha) El Chifle (131 ha) San Antonio (163 ha) La Salada (153 ha)	2005/12/02
226/86 LandSat 7 ETM 2005/12/09	El Paraíso (83 ha) Del Estado (88 ha)	2005/12/02
226/86 LandSat 7 ETM 2006/01/26	La Barrancosa (140 ha) Quilla Lauquen (94 ha) El Paraíso (77 ha) Del Estado (83 ha)	2006/01/30
224/86 LandSat 5 TM 2005/02/02	La Brava	2005/02/06

Several factors influence satellite images. Some are due to interactions of light with the atmosphere either by absorption or dispersion (Rayleigh and MIE among others), while other factors, such as distance from Sun to Earth and solar incidence angle, vary seasonally. The corresponding radiometric corrections were carried on using the conversion to radiance methods proposed by the USGS ([35] [21] [3] [4]).

In the area under study, no appropriate atmospheric data were available to accurately estimate the optical thickness of the atmosphere. Simple atmospheric correction methods, such as the dark object subtraction method to estimate haze, were not applicable because of the lack of dark objects such as deep water-bodies. To obtain surface reflectance values free from molecular scattering that reach the sensors

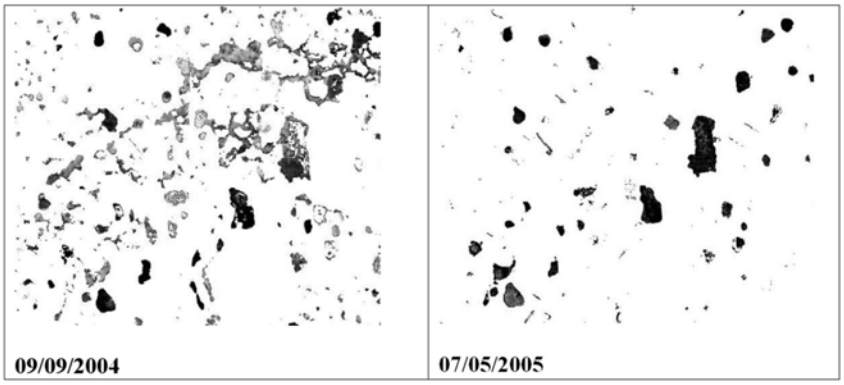
as a result of the interaction between electromagnetic radiation and the molecular components of the atmosphere, an atmospheric correction (Rayleigh method on the radiance as proposed by Kaufman [13]) was performed.

TABLE 4: Secchi disk depth (cm), chlorophyll-a concentrations (mg/m^3), and total solids concentrations (mg/l), measured in the field and computed from the satellite images using an ANN, at given shallow lakes and dates.

Shallow Lake	Date	Secchi disk depth [cm]	Chlorophyll measured [mg/m^3]	Chlorophyll computed [mg/m^3]	TS measured [mg/l]	TS computed [mg/l]
Del Estado	2005/03/04	10	325.0	425.7	1788	2051
	2005/05/07	9	600.8	575.1	2522	2372
	2005/06/08	10	667.0	679.7	2347	2597
	2005/09/12	9	533.3	554.1	2653	2327
	2005/10/14	15.5	671.2	626.7	2881	2483
El Chifle	2005/04/22	18	209.3	153.5	1756	1455
	2005/07/11	15	120.9	193.2	1942	1550
	2005/09/13	23	249.9	192.5	1736	1549
	2005/10/15	7	222.2	112.6	1953	1373
El Paraíso	2005/03/04	11	340.4	313.3	1498	1638
	2005/05/07	15	384.5	334.7	1706	1855
	2005/06/08	12	369.0	373.9	1686	1939
	2005/09/12	17	371.9	274.8	1700	1726
	2005/10/14	15	303.5	309.5	1830	1801
La Barrancosa	2005/03/04	27	89.1	178.7	1474	1519
	2005/03/13	19	155.4	147.4	1565	1451
	2005/04/22	22	95.7	138	1636	1423
	2005/07/11	18	153.5	150.4	1629	1458
	2005/09/12	22	115.2	144.5	1624	1434
	2005/10/14	17	93.2	135.5	1659	1426
	2006/01/26	11.5	71.3	103.4	1327	1234
La Brava	2005/02/02	37	45.3	46.5	507	501
La Salada	2005/04/22	23	133.9	112.3	1184	1252
	2005/07/11	17	260.2	132.4	1251	1420
	2005/09/13	16	112.0	134.2	1284	1423
	2005/10/15	13	130.6	113.9	1246	1379
Quilla Lauquen	2005/06/08	10	122.1	141.5	806	1439
	2005/10/14	22	76.3	80.5	812	788
San Antonio	2005/03/13	18	50.4	142.7	1296	1441
	2005/04/22	35	65.1	118.9	1376	1341
	2005/07/11	36	179.8	148.2	1397	1454
	2005/09/13	21	94.0	125.3	1418	1404
	2005/10/15	21	117.5	102.4	1396	1346

After the images were corrected and geo-referenced, the water bodies were separated from the rest of the image so as to avoid the influence of neighboring pixels (Map 2). Band 4 of TM was chosen for classifying inland shallow lakes using remote sensing techniques, yielding more satisfactory results than Band 5 (Priscilla Minotti,

pers. comm.). The threshold values for a water mask are to be obtained interactively. In this case, the upper threshold used for trimming the shallow lakes was 0.1. Then, an area of interest (AOI) was defined for each water body individually.



MAP 2: Examples of water masks in the same region at different dates. Note variations in water-covered areas.

Once the water-covered surfaces were isolated using the appropriate mask, their individual spectral signatures were retrieved from the average values of the set of pixels in the AOI, as shown in Figure 1. A total sample of thirty-three spectral signatures was obtained.

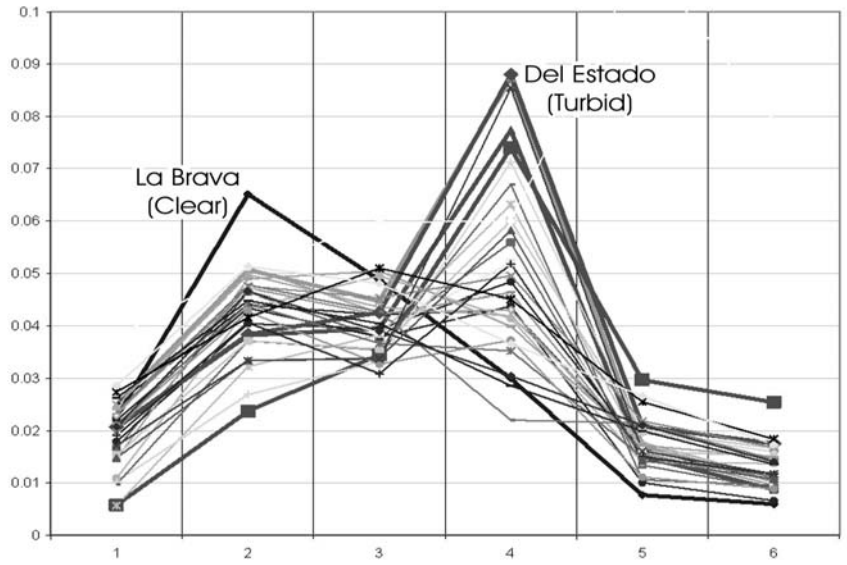


FIGURE 1: Spectral signatures of the eight selected shallow lakes recorded at different dates: (La Barrancosa (LB), Quilla Lauquen (QL), San Antonio (SA), La Salada (LS), El chifle (ECH), Del Estado (DE), La Brava (LBR) y El Paraíso (EP)). X-axis: bands (1,2,3,4,5 and 7), Y-axis: reflectance (%)

In most shallow lakes, light is predominantly absorbed by particulate matter such as phytoplankton cells and suspended sediment particles. Given that this absorbent

matter is colored, light absorption is not even for all wavelengths and some colors will penetrate more deeply than others. The apparent color of water is a very valuable characteristic for the functional differentiation of water bodies in a region. Suspended and dissolved matters contribute differently to dispersion and absorption of light [15]. For example, suspended clay particles produce greater dispersion, while dissolved organic substances produce absorption only. Phytoplankton contribute to both dispersion and absorption. Total absorption and dispersion coefficients in any given wave length are a result of the sum of the individual contribution of water, phytoplankton, suspended sediments and organic compounds [23].

Hence, it is important to detect the patterns of "extreme" spectral signatures in the sample. In Figure 1, the spectral signature drawn in the thick solid black line corresponds to "La Brava" shallow lake, which is considered a clear water body and exhibits minimum values for chlorophyll and suspended solids as well as maximum depth for Secchi disk (Class 1). The spectral signature drawn in the thick solid gray line corresponds to "Del Estado" shallow lake, a turbid water body, in extreme condition of maximum concentration of chlorophyll and suspended solids and minimum depth values for Secchi disk (Class 4). Note that the wider ranges of differences in the samples appear in Band 2 (visible spectrum) and Band 4 (near infrared).

2.4. **Canonical correlation analysis.** Canonical correlation analysis seeks to identify and quantify the associations between two set of variables [12]. This method focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The idea is first to determine the pair of linear combinations having the largest correlation. Next, to determine the pair of linear combinations having the largest correlations among all pairs uncorrelated with the initially selected pair. The process continues.

The pairs of lineal combinations are called the canonical variables, and their correlations are called canonical correlation.

When performing the analysis of covariance structure, the interest is focused in finding measures of association between two groups of variables. In this work, we are interested in identifying the associations between groups of bands. The maximization aspect of this technique is centered in reducing the problem's dimension.

Through canonical correlation analysis we observe the redundancy among the two groups which allows asserting that all the information of a group of variables is explained by other.

2.5. Artificial neural network models.

2.5.1. *ANN models vs. linear regression models.* In applied research it is very common to find situations in which a response variable should be estimated or its behavior predicted as a function of one or several predictor variables. When the response variable is quantitative the situation is considered an estimation or prediction problem. When the response variable is qualitative or categorical it is considered a classification problem.

Traditionally, the solution of these problems has been found through the use of statistical regression models such as simple or multiple regressions for prediction problems, and discriminant analysis or logistic regression models for classification problems. However, almost all these techniques are based on the so called multiple linear regression model.

This type of statistical model has been widely analyzed in the literature. Their applicability requires that the conditions on normality and equality of variances are satisfied, as well as independence, linearity, and sufficient sample size. The significance is optimal when the correlation between the predictor variables and the response is high while the correlation among predictor variables is low. In other words, colineality should not exist.

Ecological data in general do not fulfill the assumptions on normality and homogeneity of the variance. In several cases, it has been corroborated that observed environmental variables exhibit multicolineality. Hence, multiple linear regression models are not appropriate for analyzing the data involved in this study.

Progressively, artificial neural network models are being used as prediction and classification tools. In fact, ANNs have been conceptualized as nonparametric statistical techniques because they do not require the fulfillment of the theoretical assumptions of parametrical statistics. They are also considered nonlinear regression techniques.

2.5.2. *Construction of an ANN model.* An ANN model imitates the physical process of learning in the human brain. The model is formed by artificial neurons that emulate biological neurons and the synaptic connections among them, regulating them through the process of problem solving. They are appropriate for dealing with a large set of variables and their non linearity is convenient for analyzing complex systems. Once the system of neurons has been trained, the network allows the processing of imprecise information, the generalization of known responses to new situations, and the prediction of outcomes.

The network needs to be "trained" with a sufficiently large number of examples in order to be able to make the appropriate inferences. Hence, it is given groups of input data together with the expected output data (Fig. 2).

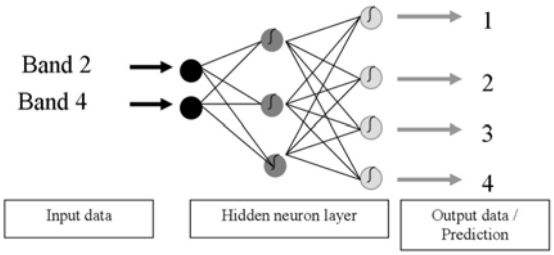


FIGURE 2: Artificial Neural Network model used for data analysis.

The links with the neurons located in the so-called hidden neuron layer then take different weights and are educated depending on the required output, thus modeling complex relationships among variables. The system requires *feedforward* and *backpropagation* processes to allow the network to get trained. The visualization of this stage is accomplished through error analysis. If the error becomes smaller and asymptotic, the network will be ready to receive new input data and predict output.

The ANN models used in this study are of the *multilayer perceptron ANN* type. ([28] [19]). The architecture is as shown in Figure 2. In each case, the training of the proposed network was performed with a *back propagation* algorithm which is a supervised learning procedure ([29] [30] [18]). It usses a method of descent on the

gradient for minimizing the global quadratic error of the output calculated by the network.

Different ANN models were built for the following objectives:

Classification of shallow lakes. A first ANN was constructed with the purpose of classifying the shallow lakes into the four proposed classes using remote sensors data. This ANN was educated using as input data Bands 2 and 4 of LandSat 5 TM and LandSat 7 ETM+ sensors, and the classes obtained from field data as output values.

Concentration of total solids and chlorophyll. The purpose is to determine from satellite data the concentration of total suspended solid matter and of chlorophyll-a, with an acceptable error margin. Thus two ANN were constructed and fed with the reflectance values retrieved from Bands 2 and 4 of LandSat 5 and LandSat 7 satellites, taking the concentration values from field samples. In the first case, both sets of data (chlorophyll-a and total solids) were run together, and in the second, they were treated separately in order to test the robustness of the method.

3. **Results.**

3.1. **Classification of shallow lakes using an ANN.**

As mentioned above, an ANN was constructed for classifying the shallow lakes into the four proposed classes. Following the classification of shallow lakes using remote sensors data, this ANN was educated using as input data Bands 2 and 4 of LandSat 5 TM and LandSat 7 ETM+ sensors and the classes obtained from field data as output.

When performing the analysis of covariance structure, two set of bands were observed: a first group included Bands 2 and 4 and a second group Bands 1, 3, 5 and 7. This was corroborated with canonical correlation analysis ($p<0.05$) (see Table 5)

TABLE 5: Canonical analysis summary regarding bands selection.

	Group 1	Group 2
No. of variables	2	4
Variance extracted	100.000%	38.2403%
Total redundancy	44.4115%	28.3240%
Variables	Left Set	Right Set
	B2, B4	B1, B3, B5, B7
Canonical R: .93267. $Chi^2(8)=88.117$ $p<0.05$		

Here the total redundancy is less than 50% in both left and right sets. Hence, both sets explain the same information. The first group was selected because bands 2 and 4 represent information adequate for the development of the *spectral unmixing* tool for deriving sub-pixel information for more accurate estimates [7], which is also part of the complete study.

The network was tested with different numbers of learning stages or epochs, different learning rates, and different numbers of neurons in the hidden layer. The model run using 80% of data for training and 20% for predicting, yield the approximation shown in Figure 3, where the plots showing both real data and the output of the ANN model coincide. In all cases the learning error becomes null in less than 20,000 epochs (Fig. 4). Hence the prediction error is zero, which means that the classification method is accurate.

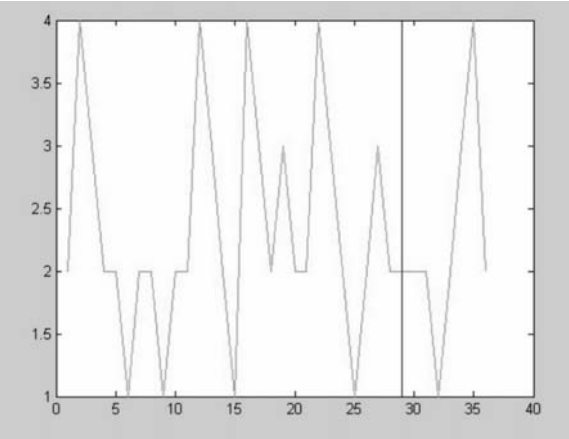


FIGURE 3: Classification of shallow lakes following their turbidity using field data (black line): the ANN model output (gray line), obtained in the training and prediction stage using 80% of data for training and 20% data for predicting, coincides. The vertical line separates both sets. X-axis: shallow lakes, Y-axis: class.

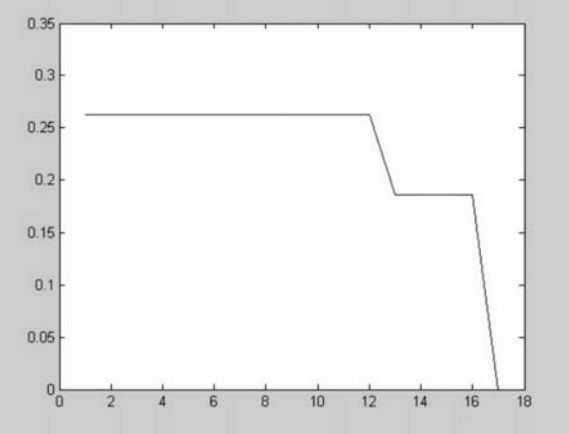


FIGURE 4: Learning error for the classification ANN model. X-axis: epochs, Y-axis: maximum absolute error.

3.2. **Concentration of total solids and chlorophyll-a using an ANN.** The networks were trained using 36 input data. As shown in Table 4, 33 values were obtained from field sampling matching the satellite image dates. In this data set, Class 1 (clear shallow lakes) is underrepresented. To balance the classes, the technique of repeating a particular datum was used. In this case, the datum corresponding to La Brava, a typical clear water shallow lake, was taken three more times. Note that both the minimum and the maximum values in the data set should be included in the learning set. Thus these two extremes were included in this set, which was then completed by a random choice of the other data in order to fulfill the portions of 75% for learning and the remaining 25% for predicting.

The first tested ANN was constructed using as output both chlorophyll-a and total solids (TS) concentrations. This model ran over 2,000,000 epochs with a learning rate of 0.00001. The number of hidden neurons varied until a best configuration was attained. The results are plotted together with the measured total solids values obtained in the field (Fig. 5).

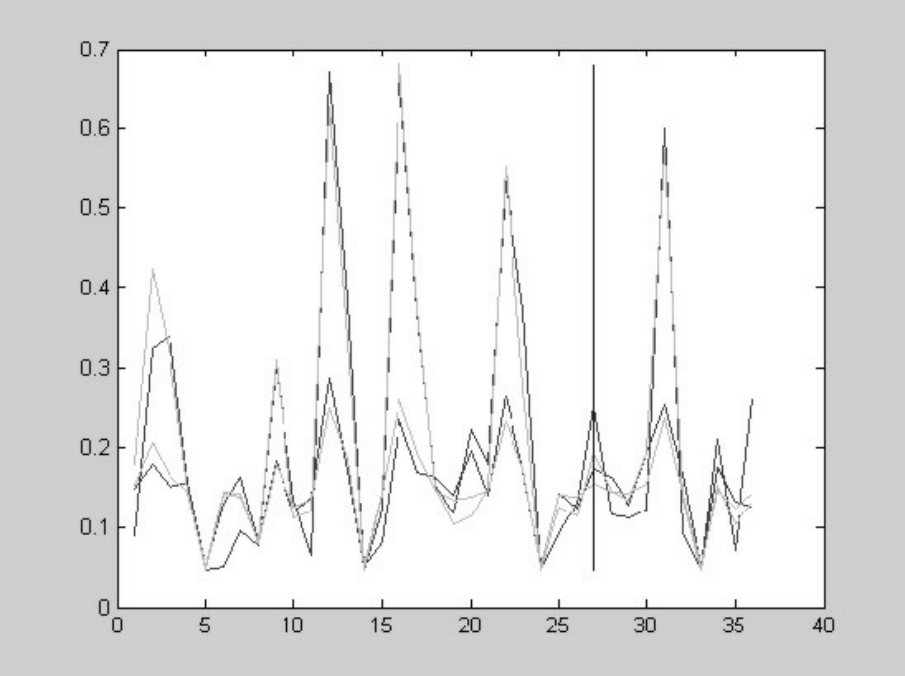
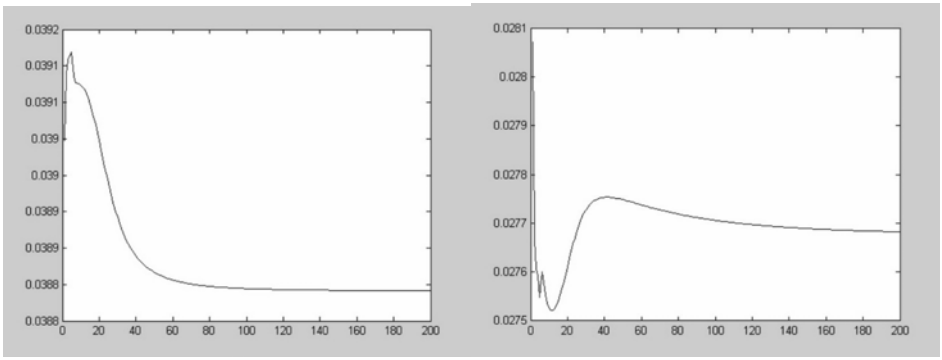


FIGURE 5:Concentration of total solids and chlorophyll-a when ANN runs are performed using both sets together, using field data (black line) and the ANN model output (gray line) obtained in the training and prediction stage using 75% of data for training and 25% data for predicting. The vertical line separates both sets. X-axis: shallow lakes, Y-axis: chlorophyll-a [10^{-3} mg/m^3] and total solids [10^{-4} mg/l] concentrations.

Learning and predicting errors expressed as the maximum module of the absolute differences and divided by 1,000 (chlorophyll-a) y 10,000 (TS) behave asymptotically (Fig. 6 (a) and (b)). The learning error exhibits a decreasing tendency and yielded the results for chlorophyll and TS shown in Table 4, columns 5 and 7.

Separate runs were performed to observe the behavior of an ANN for each of the data sets. For chlorophyll-a, ANN models with two, three and more neurons in the hidden layer were run over 1, 1.5 and 2 million epochs with learning rates between 0.00001 and 0.0000001. The best results were obtained with two neurons in the hidden layer and a learning rate of 0.00001 (Fig. 7).



(a) (b)

FIGURE 6. Learning (a) and predicting (b) errors for the concentration of chlorophyll and total solids when ANN runs are performed with both data sets jointly. X-axis: epochs, Y-axis: maximum absolute error.

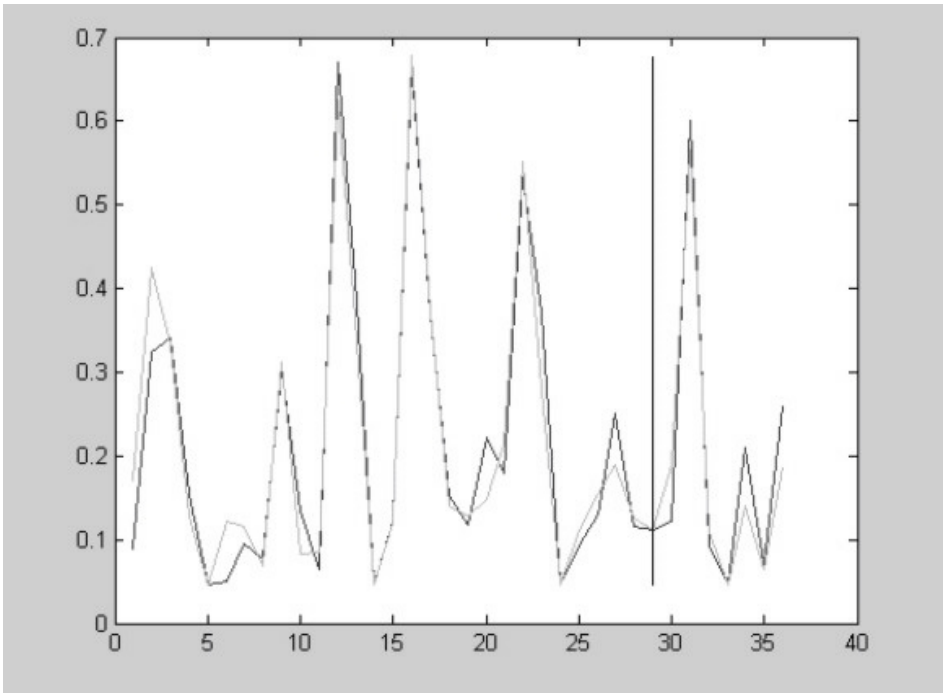
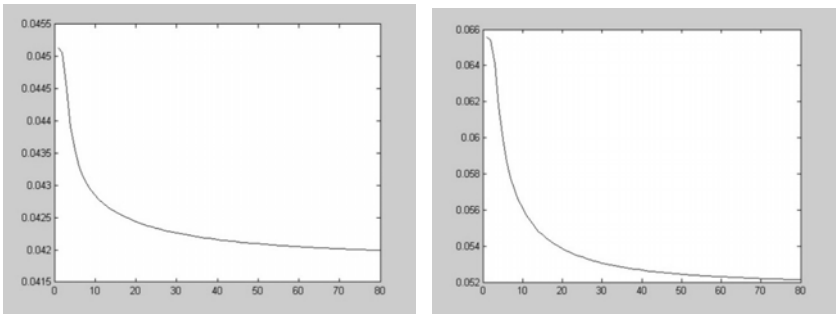


FIGURE 7: Concentration of chlorophyll-a using field data (black line) and the ANN model output (gray line) obtained in the training and prediction stage using 80% of data for training and 20% data for predicting. The vertical line separates both sets. X-axis: shallow lakes, Y-axis: chlorophyll concentration [10^{-3} mg/m^3].

Learning and predicting errors were approximately 42 and 52 mg/m^3 respectively (Fig. 8 (a) and (b)).



(a) (b)

FIGURE 8. Learning (a) and predicting (b) errors for the concentration of total solids. X-axis: epochs, Y-axis: maximum absolute error.

When the hidden layer contained three neurons, the learning error was reduced but the prediction error increased monotonely. Increasing the number of neurons does not lead to convergence in the results.

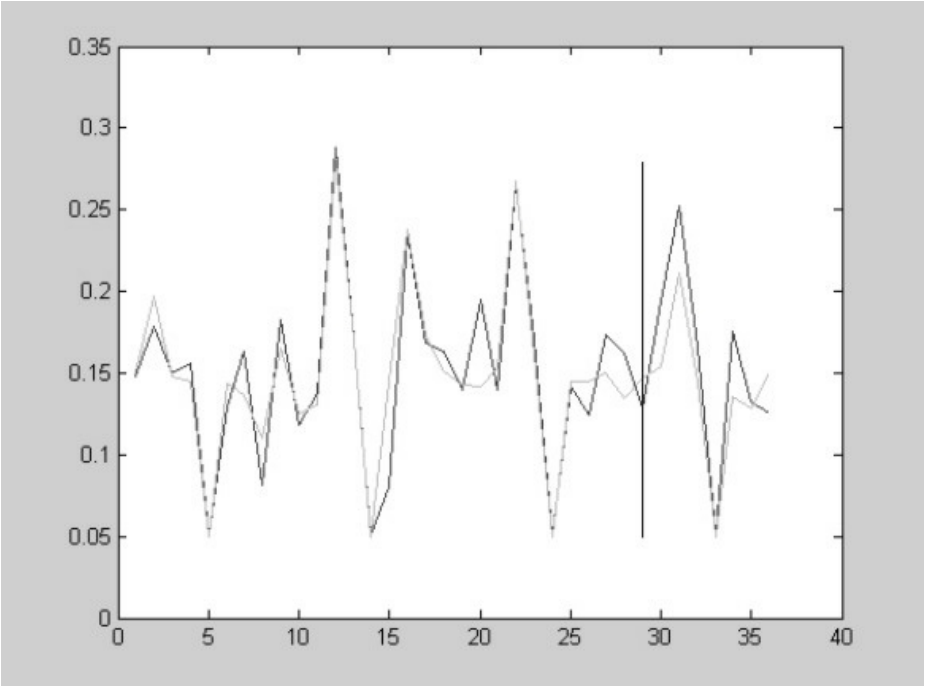
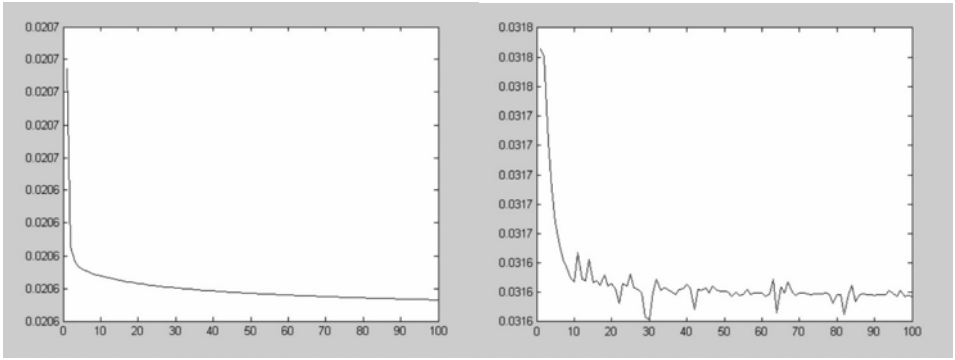


FIGURE 9: Concentration of total solids (10^{-3} mg/l) using field data (black line) and the ANN model output (gray line) obtained in the training and prediction stage using 80% of data for training and 20% data for predicting. The vertical line separates both sets. X-axis: shallow lakes, Y-axis: total solids concentration [10^{-4} mg/l].

The same analysis was performed for TS. Similar results were obtained for two and three neurons in the hidden layer and a learning rate of 0.00001 (Fig. 9).

Learning and predicting errors were approximately 200 and 320 mg/l respectively (Fig. 10 (a) and (b)).



(a) (b)

FIGURE 10. Learning (a) and predicting (b) errors for the concentration of total solids X-axis: epochs, Y-axis: maximum absolute error.

To test the performance of the method, both ANN models were run taking the data of all but one shallow lake for training and then predicting the values for the remaining one. This allowed to estimate how well did the ANN model respond when attempting to predict variability within the same water body. Table 6 shows the results when chlorophyll-a and total solids values are within a reasonable range from the median values of the samples.

TABLE 6: Performance of the ANN models as predictors of TS and chlorophyll-a concentrations for different target lakes: a random selection of shallow lakes, La Barrancosa, San Antonio, El Chifle-La Salada and El Chifle-El Paraiso shallow lakes.

	Random Sample	La Barrancosa	San Antonio	Chifle – La Salada	Chifle – El Paraiso
Median Total Solids	1624 mg/l	1624 mg/l	1396 mg/l	1849-1249 mg/l	1849-1700 mg/l
Absolute learning error	200 mg/l	224 mg/l	216 mg/l	170 mg/l	156 mg/l
Absolute predicting error	320 mg/l	205 mg/l	217 mg/l	357 mg/l	821 mg/l
Mean square error	0.01	0.03	0.02	0.03	0.04
Determination coefficient	0.94	0.87	0.87	0.87	0.79

	Random Sample	La Barrancosa	San Antonio	Chifle – La Salada	El Paraiso
Median chlorophyll-a	153.5 mg/ m^3	95.7 mg/m^3	94.0 mg/m^3	216-132 Mg/m^3	369 mg/m3
Absolute learning error	42 mg/m^3	42 mg/m^3	42 mg/m^3	37 mg/m^3	45 mg/m^3
Absolute predicting error	52 mg/m^3	48 mg/m^3	63 mg/m^3	74 mg/m^3	182 mg/m^3
Mean square error	0.09	0.08	0.08	0.09	0.16
Determination coefficient	0.86	0.89	0.88	087	0.77

3.3. **Validation of results.** To validate the results obtained from the ANN models, a dispersion diagram was done. It allows analyzing the behavior of measured TS and computed TS from available data, excluding the values that distort the sample.

Figure 11 shows a positive tendency and a strong association between the two sets of variables. A simple linear regression analysis, taking as response variable the measured TS values, was performed with the purpose of quantifying the tendency. The regression analysis for TS yields a determination coefficient $R^2 = 0.94$, a mean square error of 0.01 and a fitted regression line

$$Measured \ln TS = 0.41 + 0.94 \, Computed \ln TS \ (p < 0.01).$$

The graph, confidence bands, and prediction are shown in Figure 11.

Regarding measured and computed chlorophyll-a concentrations, the regression analysis yields a determination coefficient $R^2 = 0.86$, a mean square error of 0.09 and a fitted regression line.

$$Measured \ln CL = -0.09 + 1.01 \, Computed \ln CL \ (p < 0.01).$$

The graph, confidence bands, and prediction are shown in Figure 12.

In both cases, the residuals analyses indicate independence, normality, and homogeneity in the variance, which permits to assert that the regression model is a good fit. The same simple linear regression analysis was repeated on the measured and the calculated data obtained from each ANN model. Table 6 synthesizes the results for different ANN output, when each of the shallow lakes in was predicted using the remaining lakes data for training the network.

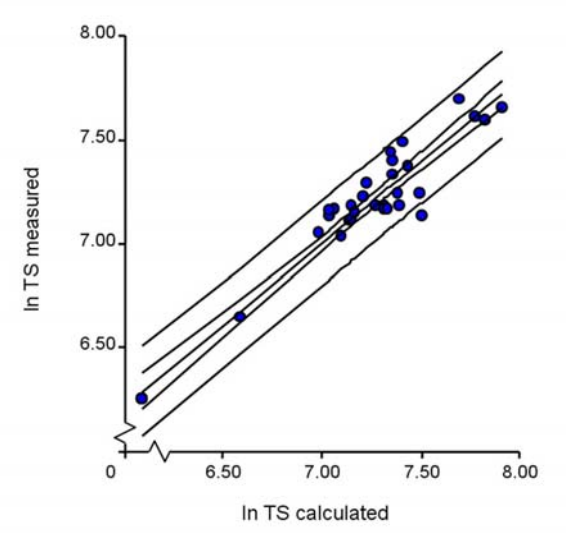


FIGURE 11: Regression analysis of measured and calculated total solids with confidence and prediction bands.

When the median value of the selected shallow lake is close to the total sample median, a positive association (> 85%) between the two sets (calculated and measured values) can be observed. The values of the absolute predicting error confirm the good fit. When the median value of the selected shallow lake is close to an extreme value, the determination coefficient is lower and the values of the absolute predicting error are high.

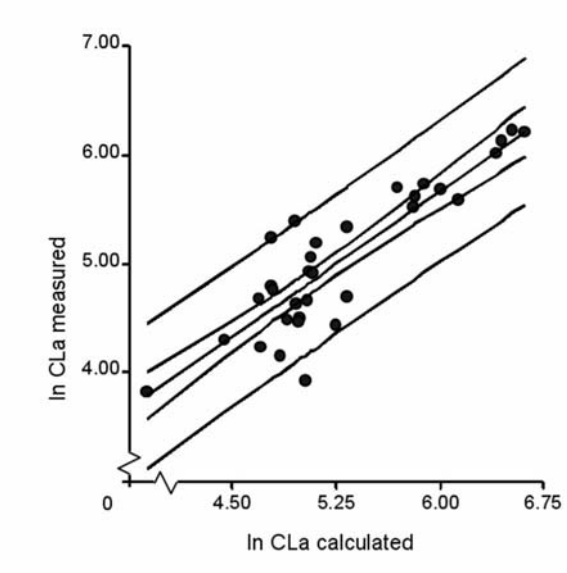


FIGURE 12: Regression analysis of measured and calculated chlorophyll with confidence and prediction bands.

4. **Discussion and conclusions.** Pampean shallow lakes are mostly eutrophic and hypereutrophic. Our collection of samples shows chlorophyll-a values between 45 and 1400 mg/m^3 and total solids values between 507 and 2881 mg/l. These lakes challenge the usual methodologies because of the influence that this characteristic has on the optic properties captured by the satellite sensors. Not only they exhibit high concentrations of chlorophyll-a and total solids but also high temporal variability within the same water body.

Much of the published research on shallow lakes is based on LandSat imagery, and the methods have been mostly developed for water bodies that exhibit chlorophyll-a concentrations below 100 mg/m^3 ([34] [33] [16]).

Our ANN models have proven to be very adequate for the classification of shallow lakes following their turbidity, yielding accurate output relative to the quantile probability distribution obtained from field data. Thus this methodology can be confidently used for classifying other shallow lakes of the region.

The combination of field data, spectral determination and ANN models for estimating total solids and chlorophyll-a concentrations from bands 2 and 4 of Landsat TM and ETM+ on the pampean shallow lakes produced satisfactory results. ANN models here developed seem capable of dealing with the above-mentioned wide range of concentrations with low error and good confidence values, as shown in Table 6. This is stressed by the validation using regression analysis. The predicted values here obtained validate the use of ANN models for concentrations of chlorophyll-a and total solids as high as 260 mg/m^3 and 1950 mg/l respectively.

Even though the training set is relatively small, our ANN models yield very good predictions when the predicted values fall within a certain range of the median values of the training data set. It can be expected that the precision of the prediction will improve as the training set becomes larger. The errors increase when the predicted values are closer to the extreme values of the training data set. This is

very reasonable given that there are fewer input data points in the neighborhood of the extremes. However, the determination coefficients thus obtained confirm that the results from ANN models are very satisfactory.

This method can be used to gain knowledge on the trophic state of the numerous shallow lakes in this region in real time from Bands 2 and 4 of the LandSat TM and ETM+ images. ANN models here constructed are useful because they permit one to extrapolate the information obtained in field work in a number of shallow lakes to the whole set of similar water bodies in the region and thus contribute to understanding their dynamics as a system. It is important that the database used for training the ANN models includes a balanced representation of typical shallow lakes in all the classes.

The methodology presented here is one initial step in the construction of a database that will allow the study of patterns of behavior and responses of these shallow lakes to different phenomena. From the optical properties of water bodies, it is possible to quantify the components that influence these properties. Artificial neural network models appear to be efficient tools for estimating the contribution of particular components, such as chlorophyll-a and total solids. The temporal analysis of these variables together with the frequencies of appearance of temporary shallow lakes, recurrence of flooded areas, changes in individual lake area, spatial behavior of permanent water bodies will contribute to understand the complex dynamics of the system. Hence it is possible to use remote sensors data for monitoring the evolution at either local or regional scale of these variables, or to infer the changes in state or stages of shallow lakes by following changes in turbidity.

Acknowledgements. We thank the National Commission for Space Activities (CONAE) for providing LandSat images. Field work was conducted by Fabian Grosman and Pablo Sanzano while Oscar Diaz and Viviana Colasurdo were responsible of laboratory analyses. They are all members of the Multidisciplinary Institute on Ecosystems and Sustainable Development. Federico Dukatz has a Graduate Studies Scholarship from the Scientific Research Commission of the Province of Buenos Aires (CICPBA). We thank Priscilla Minotti for valuable suggestions on image processing. We also thank the referees for their valuable comments and suggestions.

This work is dedicated to Prof. Dr. Tom G. Hallam on his 70th birthday.

REFERENCES

[1] R. Aguirre-Gomez, *Detection of total suspended sediments in the North Sea using AVHRR and ship data*, International Journal of Remote Sensing, **21** (2000), 1583-1596.

[2] S. Bagheri, C. Zetlin, and R. Dios, *Estimation of optical properties of nearshore water,* International Journal of Remote Sensing, **20** (1999), 3393-3397.

[3] G. Chander and B. L. Markham, *Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges,* IEEE Transactions on Geoscience and Remote Sensing, **41** (2003), 2674–2677.

[4] G. Chander, B. L. Markham, and J. A. Barsi, *Revised Landsat-5 Thematic Mapper Radiometric Calibration*, IEEE Geoscience and Remote Sensing Letters, **4** (2007), 490-494.

[5] N.V. Dangavs, *Los ambientes lenticos de la Pampasia Bonaerense, Republica Argentina*, in "Agua Problematica Regional: Enfoques y Perspectivas en el Aprovechamiento de Recursos Hidricos", EUDEBA, Buenos Aires, 1998.

[6] F. Dukatz and R. Ferrati, *Sistematizacion del analisis y clasificacion de cuerpos de agua segun permanencia mediante sensores remotos*, in: "International Congress on Development, Environment and Natural Resources: Sustainability at Multiple Levels and Scales", (eds. J. Feyen, L.F. Aguirre, and M. Moraes R.), Cochabamba, Bolivia, **II** (2007), 1102-1107.

[7] R. Ferrati, C. Marinelli, F. Dukatz, R. Cepeda, and G. Canziani, *Categorizacion de cuerpos de agua segun sus propiedades opticas*, in: "International Congress on Development, Environment and Natural Resources: Sustainability at Multiple Levels and Scales", (eds. J. Feyen, L.F. Aguirre, and M. Moraes R.), Cochabamba, Bolivia, **II** (2007), 1108-1115.
[8] J. Frenguelli, *Rasgos generales de la hidrografía de la Provincia de Buenos Aires,* LEMIT II **62** (1956), 1-19.
[9] F. Grosman, P. Sanzano, O. Diaz, V. Colasurdo, R. Ferrati, F. Dukatz, and G. Canziani, *Lagunas de la region pampeana de Argentina: estructura y funcionamiento a partir de aspectos sociales, economicos, culturales y naturales,* in: "International Congress on Development, Environment and Natural Resources: Sustainability at Multiple Levels and Scales", (eds. J. Feyen, L.F. Aguirre, and M. Moraes R.), Cochabamba, Bolivia. **II** (2007), 1130-1137.
[10] J. A. Harrington Jr, E. R. Schiebe and J. E. Nix, *Remote sensing of Lake Chicot, Arkansas: monitoring suspended sediments, turbidity, and Secchi depth with Landsat MSS data,* Remote Sensing of Environment, **39** (1992), 15-27.
[11] S. J. Hudson, G. F. Moore, A. J. Bale, K. R. Dywer, and J. Aitken, *An operational approach to determining suspended sediment distributions in the Humber estuary by airborne multispectral imagery*, in: "Proceedings of the First Airborne Remote Sensing Conference and Exhibition", Strasburg, France, 1994.
[12] R. Johnson and D. Wichern, "Applied Multivariate Statistical Analysis", Prentice Hall, New Jersey, 1992.
[13] Y. J. Kaufman, *The atmospheric effect on remote sensing and its correction*, in: "Theory and Application of Optical Remote Sensing", (ed. G. Asrar), Wiley, New York, 1989.
[14] L. E. Keiner and X. H. Yah, *A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery*, Remote Sensing of Environment, **66** (1998), 153-165.
[15] J. T. O. Kirk, "Light and Photosynthesis in Aquatic Ecosystems", Cambridge University Press, Cambridge, 1994.
[16] Y. Liu, M. A. Islam, and J. Gao, *Quantification of shallow water quality parameters by means of remote sensing*, Progress in Physical Geography, **27** (2003), 24-43.
[17] M. J. Lopez-Garcia and V. Caselles, *Use of Thematic Mapper data to assess water quality in Albufera Lagoon of Valencia (Spain)*, in "Advances in Digital Image Processing, Proceedengs 13th Annual Conference of the Remote Sensing Society", Nottingham, (1987), 510-519.
[18] J. L. McClelland and D. E. Rumelhart, "Explorations in Parallel Distributed Processing", MIT Press, Cambridge, 1988.
[19] M. L. Minsky and S. A. Papert, "Perceptrons, Expanded Edition", MIT Press, Cambridge, 1969.
[20] G. K. Moore, *Satellite remote sensing of water turbidity*, Hydrological Sciences Bulletin, **25** (1980), 407-421.
[21] NASA/USGS LandSat 7 Program, "LandSat 7 Science Data Users Handbook", 1998, http://landsathandbook.gsfc.nasa.gov/handbook.html
[22] L. G. Olmanson, M. E. Bauer, and P. L. Brezonik, *Use of Landsat imagery to develop a Water Quality atlas of Minnesota's 10,000 Lakes.* in: "Pecora 15/Land Satellite Information IV Conference/ISPRS Commission I Midterm Symposium/FIEOS 2002 Conference Proceedings", 2002. http://www.isprs.org/commission1/proceedings02/paper/00102.pdf
[23] L. Prieur and S. Sathyendranath, *An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter and other particular materials*, Limnology and Oceanography, **26** (1981), 671-689.
[24] R. Quiros, A. Renella, M. Boveri, J. Rosso, and A. Sosnovsky, *Factores que afectan la estructura y funcionamiento de las lagunas pampeanas,* Ecología Austral, **12** (2002), 175-185.
[25] J. C. Rimmer, M. B. Collins and C. Pattiaratchi, *Mappìng of coastal water quality in coastal waters using airborne Thematic Mapper data.* International Journal of Remote Sensing, **8** (1987), 85-102.
[26] R. A. Ringuelet, "Ecología Acuática Continental". EUDEBA, Buenos Aires, 1962.
[27] V.K. Rohatgi, "An Introduction to Probability Theory and Mathematical Statistics". John Wiley & Sons, New York, 1976.
[28] F. Rosenblatt, "Principles of Neurodynamics", Spartan, New York, 1962.
[29] D.E. Rumelhart, G. E. Hinton, and R. J. Willams, *Learning Internal Representations by Error Propagation*, in: "Parallel Distributed Processing, Explorations in the Microestruture

of Cognition, Vol. 1: Foundations" (eds. D. E. Rumelhart and J. L. McClelland), MIT Press, Cambridge, 1986.

[30] D. E. Rumelhart, G. E. Hinton, and R. J. Willams, *Learning Representations by Back-Propagation Error*, Nature, **323** (1986), 533-536.

[31] M. Scheffer, "Ecology of Shallow Lakes", Kluwer Academic Publishers, Dordrecht, 2004.

[32] H. Shimoda, M. Etaya, T. Sakata, L. Goda, and K. Stelczes, *Water quality monitoring of Lake Balaton using Landsat MSS data*, in "Remote Sensing for Resources Development and Environmental Management. Proceedings 7th ISPRS Commission VII Symposiumj", Enschede, Rotterdam, **2** (1986), 765-770.

[33] E. Svab, A. N. Tyler, T. Preston, M. Presing, and K. Balogh, *Characterizing the spectral reflectance of algae in lake waters with high suspended sediment concentrations,* International Journal of Remote Sensing, **26** (2005), 919-928.

[34] S. Thiemann and H. Kaufmann, *Determination of chlorophyll content and trophic state of lakes using field spectrometer and IRS-1C satellite data in the Mecklenburg lake district, Germany,* Remote Sensing of Environment, **73** (2000), 227-235.

[35] USGS, "Multi-Resolution Land Characteristics 2001 (MRLC2001) Image Processing Procedure", 2006. http://landcover.usgs.gov/pdf/image_preprocessing.pdf

Received on December 31, 2007. Accepted on July 2, 2008.

E-mail address: canziani@exa.unicen.edu.ar
E-mail address: rferrati@exa.unicen.edu.ar
E-mail address: cmarine@exa.unicen.edu.ar
E-mail address: fdukatz@exa.unicen.edu.ar