



## Chlorophyll *a* spatial inference using artificial neural network from multispectral images and *in situ* measurements

MONIQUE S. FERREIRA<sup>1</sup> and MARIA DE LOURDES B.T. GALO<sup>2</sup>

<sup>1</sup>Programa de Pós Graduação em Ciências Cartográficas, FCT/ UNESP,  
Rua Roberto Simonsen, 305, 19060-900 Presidente Prudente, SP, Brasil

<sup>2</sup>Departamento de Cartografia, FCT/ UNESP,  
Rua Roberto Simonsen, 305, 19060-900 Presidente Prudente, SP, Brasil

*Manuscript received on april 3, 2012; accepted for publication on October 5, 2012*

### ABSTRACT

Considering the importance of monitoring the water quality parameters, remote sensing is a practicable alternative to limnological variables detection, which interacts with electromagnetic radiation, called optically active components (OAC). Among these, the phytoplankton pigment chlorophyll *a* is the most representative pigment of photosynthetic activity in all classes of algae. In this sense, this work aims to develop a method of spatial inference of chlorophyll *a* concentration using Artificial Neural Networks (ANN). To achieve this purpose, a multispectral image and fluorometric measurements were used as input data. The multispectral image was processed and the net training and validation dataset were carefully chosen. From this, the neural net architecture and its parameters were defined to model the variable of interest. In the end of training phase, the trained network was applied to the image and a qualitative analysis was done. Thus, it was noticed that the integration of fluorometric and multispectral data provided good results in the chlorophyll *a* inference, when combined in a structure of artificial neural networks.

**Key words:** remote sensing of water, fluorescence, chlorophyll *a*, spatial inference, artificial neural network.

### INTRODUCTION

Water is a natural resource of vital ecological and economic importance, essential for the survival of all life forms on the planet. Although its limitation in quality and quantity are widely reported, water resources currently available have been used over the years without adequate planning and control.

In order to evaluate and monitor water resources quality, researchers and managers analyze various substances and organisms in the water, trying to relate their presence to water quality standards.

Among these, the chlorophyll *a*, which is the primary pigment of all photosynthetic organisms and it is present in all types of algae (Wetzel 2001), constitutes approximately 1 to 2% of the dry weight of planktonic algae and has been considered an indicator of algal biomass and also the trophic state of aquatic systems (Apha 1998).

Representative measures of chlorophyll *a* concentration are fundamental to infer about the ecological processes that occur in a aquatic system, but they are difficult to achieve, particularly when the sample size should be too large to allow a suitable analysis of the aquatic system.

---

Correspondence to: Monique Sacardo Ferreira  
E-mail: mosacardo@gmail.com

In this context, as cited by Londe (2008), remote sensing is an effective alternative for monitoring water quality, for enabling studies in appropriate temporal and spatial scales. Moreover, chlorophyll *a* is considered an optically active component (OAC) due to the property of interact with electromagnetic radiation, causing different patterns of absorption and scattering. Thus, it is possible correlate its presence to reflectance measurements obtained from spectral bands of remote sensing images.

Also, chlorophyll is considered a fluorescent substance, *i.e.*, when it is irradiated at a specific wavelength (excitation), the electronic state of the molecule changes and it tends to return quickly to ground state by a process of radioactive decay at a wavelength typically longer than the excitation (Holler et al. 2009).

According to Barbosa (2003), monitoring chlorophyll fluorescence to obtain information about the aquatic system is a valid concept, since fluorescence is measured externally to the photosynthetic organism and can be detected in a non-invasively and non-destructive way. Therewithal, remote sensors can be used with this intention, with the additional advantage that the measures can be taken at different scales, both spatial (from microns to kilometers) and temporal (from microseconds to months).

Thereby, the spectral peculiarities of chlorophyll provides several ways for its detection in large-scale in aquatic systems. Scientific community should study methods for integrating data sources about the variable. If the purpose is the knowledge of the spatial distribution of the pigment, inference techniques should be used. Among these techniques, the use of Artificial Neural Networks (ANN) can be considered a powerful tool for spatial inference of environmental variables using remote sensing data.

As stated by Haykin (1999), an ANN is a parallel distributed processor made up of simple processing units, which have a natural propensity

for storing experiential knowledge and making it available for use. It resembles the human brain in two aspects: (i) the network acquires knowledge using a learning process, and (ii) connection strengths between neurons, known as synaptic weights, are used to store the acquired knowledge.

The increasing acceptance of artificial neural networks in remote sensing due to its ability to (i) learn complex patterns, considering any complex non-linear relationship between the dependent and explanatory variables, (ii) generalization in noisy environments, which makes robust ANNs solutions in the presence of incomplete or inaccurate data, (iii) incorporate *a priori* knowledge and real restrictions in data analysis, (iv) to integrate different types of data for analysis, due to the absence of assumptions about the data set (Mas and Flores 2008).

The inference of biophysical parameters from remote sensing data using artificial neural networks assumes that the remote sensing data serve as input data to the network. A training sample, selected from the input data, is presented to net. After the training and definition of a structure of weight connections, the neural net is able to produce a numeric output for the studied parameter. The approaches that combine ANNs and remote sensing data have been applied to various fields of knowledge such as meteorology (Jang et al. 2004), agriculture (Danson et al. 2003), pedology (Chang and Islam 2000), oceanography (Bukton et al. 1999, Gross et al. 2000, Keiner and Brown 1999, Keiner and Yan 1998, Schiller and Doerffer 1999), among others.

Considering the above, it was identified a small water body for public water supply, located within the urban limits of Marilia city, São Paulo State, called Cascata reservoir.

The research aimed to perform the spatial inference of chlorophyll *a*, using an artificial neural network, trained from a set of georeferenced measures of chlorophyll concentration collected *in situ*, and reflectance values recorded in a multispectral image. The following specific objectives were outlined:

- To verify the suitability of the field fluorometry to register low levels of chlorophyll *a* and to enable the acquisition of a dense sampling of the variable;
- To evaluate the potential of a multispectral image with high spatial resolution in the detection of chlorophyll *a* in the aquatic system;
- To analyze the performance of an artificial neural network in the process of spatial inference.

### MATERIALS AND METHODS

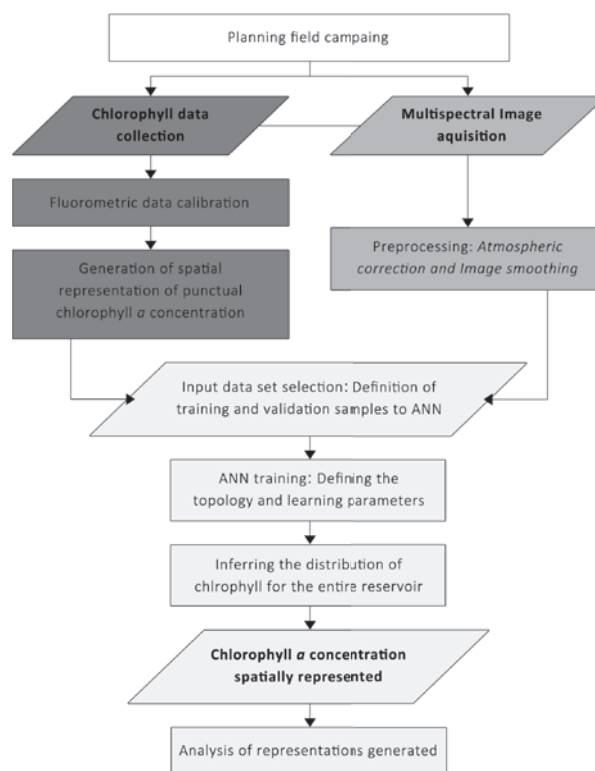
The chlorophyll *a* spatial inference was made using an artificial neural network based on georeferenced measures taken by a field fluorometer between March 17-18<sup>th</sup>, 2011, and a radiometrically corrected multispectral image taken on March 20<sup>th</sup>, 2011. The study was developed as follows:

- Field survey complemented by the acquisition of a multispectral image, obtaining the values of chlorophyll *a* in points with known coordinates;
- Worldview-2 multispectral image treatment and analysis;
- Definition of the architecture, input data set and training of the ANN;
- Application of the trained network to the image;
- Analysis of generated representations.

The sequence of these activities is presented in Figure 1, highlighting the three methodological approaches of the work.

To develop the research, different data sources were used: a multispectral image, a field fluorometer to acquire chlorophyll *a* concentration measurements, an artificial neural network simulator and other software and ancillary data.

Besides of fluorometry data, a multispectral image Worldview-2 was used as input for the



**Figure 1** - Experimental design.  
ANN: Artificial Neural Networks.

spatial inference. It is an orbital system developed by Digital Globe which is considered the successor of QuickBird series. According to the company, Worldview-2 is the first commercial sensor to provide images with high spatial resolution in eight multispectral bands. The sensor characteristics are presented in Table I.

In the atmospheric correction of the image, a MODIS (MODerate resolution Imaging Spectroradiometer) product (MOD07\_L2) was used. This product is generated from data collected by MODIS sensor installed on TERRA platform.

The acquisition of georeferenced measurements of chlorophyll *a* was performed using a field fluorometer, which quantifies the pigment through its fluorescence. The equipment used was the 10-AU, produced by Turner Designs enterprise, from United States of America. It provides direct measurements of chlorophyll *a* fluorescence and

**TABLE I**  
**Worldview-2 system specifications.**

Launch	October 8, 2009
Orbit Type	Sun-synchronous
Orbit Altitude	770 km
Spatial Resolution (at Nadir)	Panchromatic: 0,46 m Multispectral: 1,84 m
Spectral Bands	Coastal (400 – 450 nm) Blue (450 – 510 nm) Green (510 – 580 nm) Yellow (585 – 625 nm) Red (630 – 690 nm) Red Edge (705 – 745 nm) Near InfraRed 1 - NIR-1 (770 – 895 nm) Near Infra Red 2 - NIR-2 (860 – 1,040 nm)
Dynamic Range	11 bits per pixel

Font: Adapted from Digital Globe 2010.

other phytoplankton pigments in live algal cells. The technique *in vivo* fluorescence (IVF) is ideal for profiling in real time, because it does not require special treatment of the samples (Turner Designs 2004).

To locate correctly the sample elements it was used a GNSS (Global Navigation Satellite Systems) receptor.

The software used in the inference by ANN was the Java version of the Stuttgart Neural Network Simulator - JavaNNS (Fischer et al. 2001), available at <http://www.ra.cs.uni-tuebingen.de/SNNS/>. IDRISI Andes software was used to convert the net input data for use in JavaNNS and to convert the output data from it. To the image preprocessing and to select data to training the ANN, the software ENVI 4.8 was used. Finally, ArcGIS was used to produce the maps and spreadsheets to assisting the task of data conversion.

#### CHLOROPHYLL A CONCENTRATION BY FIELD FLUOROMETRY

The data of chlorophyll taken by 10-AU fluorometer are a relative measure, because the fluorescence intensity is directly proportional to the fluorophore

concentration (Holler et al. 2009). Then, it is necessary to calibrate the fluorometer measures using sample elements which the pigment are determined by a recognized laboratory method, enabling to obtain the actual concentration of chlorophyll *a*.

The field campaign carried out between March 17-18th, 2011; it aimed to acquire measures and also to calibrate the instrument. Simultaneously the acquisition of the fluorometric measures, a punctual sampling was made in the first day, in which aliquots of water were collected in each point for laboratory determination of chlorophyll *a*. The points were defined based on a random sampling scheme, and were properly positioned using a GNSS receiver Hiper GGD. The survey method used was the Stop-and-Go, or relative semi-kinematic. A base receiver was positioned on the side of the reservoir, and its coordinates were defined by relative static positioning.

The aliquots of water collected were stored at low temperature for subsequent determination of chlorophyll *a* concentration using the High Performance Liquid Chromatography technique (HPLC) in the laboratory of the Núcleo de Pesquisas Avançadas em Matologia (NUPAM / FCA) at the Universidade Estadual Paulista (UNESP) Botucatu.

The determination of chlorophyll concentration in the sample elements was necessary to do the calibration curve and consequent adjustment of the readings of the field fluorometer. To do this curve, 8 sample elements were used, ensuring a good fit.

The collection in continuous flow was carried out on March 18<sup>th</sup>, 2011. The course of collection was recorded with a GNSS receiver, this time using kinematic relative positioning method with the same base point used in Stop-and-Go positioning. To synchronize fluorescence measures and the coordinates of the points, the two instruments had their collection rates set to 5 seconds and were triggered simultaneously. The lag between them was 2 seconds and was considered acceptable. Thus, with fluorometer data calibrated

and synchronized with UTM coordinates, a dense sampling with 728 observations of the actual concentration of chlorophyll *a* was obtained in the study area.

#### MULTISPECTRAL IMAGE PREPROCESSING TO CHLOROPHYLL DETECTION

The multispectral image Worldview-2 was taken on March 20<sup>th</sup> 2011. In this case, the time lag with the field campaign was two days. Since there was no sudden weather change, or direct interventions in the water resource that could significantly alter its limnological characteristics, this delay was acceptable.

This scene was acquired at the process level 2A or Ortho Ready Standard Imagery. This is an image with basic level of radiometric and geometric correction, including correction of sensors inherent errors and relating the scene to a reference system.

Atmospheric correction is essential for handling the spectral attributes of a scene. In the present investigation the atmospheric correction was performed in ENVI 4.8 software, using the FLAASH module (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube).

First, it was performed the calibration of Worldview-2 radiance values. ITT (2010) cites that the images of this sensor are provided to users on relative radiance values, therefore the initial processing applied to spectral bands was the radiometric calibration, resulting in absolute radiance values in the sensor. The procedure for this calibration is described in Updike and Comp (2010), and is also implemented in ENVI.

Thereafter, atmospheric correction was made. For this, FLAASH requests the parameters of the geometry scene acquisition and atmospheric parameters related with the atmospheric model, water vapor concentration, aerosol model and visibility.

The atmospheric model was defined as tropical, since this is the geographical region of study area; the parameter of water vapor concentration was

based on the standard value for a tropical model, with the addition of a multiplier factor that was determined from the MOD07\_L2 product, whose data were collected simultaneously with the Worldview-2 acquisition.

The aerosol model was urban, since the reservoir is located in the urban area of Marilia city and the visibility value was empirically defined as suggested by Ennes (2008), that cites the need to estimate initially the visibility parameter and then reset it until the reflectance spectra of the resulting model are consistent with the spectral response expected for the targets.

After atmospheric correction of the scene was performed, the target of interest was isolated. This made it possible to analyze separately its spectral attributes and later modeling the occurrence of chlorophyll *a* in the study area. Water body isolation was performed using the result of an image segmentation of spectral band 7 (NIR-1 is the region that the spectral response of the water is almost null, which permit to differentiate it from other targets), whose mask was used to extract the surface reflectance values just inside the water body in each Worldview-2 spectral band.

The next steps aimed at exploring the spectral attributes of the image. Although the scene has 8 spectral bands, only 7 were used in the study. This because the NIR-2 band (860-104 nm) is not a region suitable for spectral studies of water, since this target absorbs all incident electromagnetic radiation.

A visual inspection of the image shows a speckled effect throughout the water body. Considering the fact that significant changes in the spectral response of a small water body rarely occur abruptly and since the spatial resolution of the image is 2 meters, this effect may not be associated with the response of water itself.

In this sense, the use of filtering techniques to remove the effect of spectral variation, since this does not represent the actual interaction of REM in the water column, was the option found to match the



use of these images to infer chlorophyll. For image smoothing, is commonly used low-pass filters, which preserve the low frequency components of an image, smoothing it out. Richards and Jia (2006) highlighted that the low-pass filters based on the mean and median are the most commonly used smoothing method. While the mean filter smoothes the image in general and also can remove edges, the median filter preserves it, and may even enhance this information. In this sense, searching for a product suitable for spatial inference, it was performed a smoothing with two types of filters, both using a neighborhood of 3 x 3 pixels.

#### ARTIFICIAL NEURAL NETWORK TO SPATIAL INFERENCE OF CHLOROPHYLL-*a*

To the geographical inference process it is necessary to define the input and output data to insert to the network. Thus, the Worldview-2 spectral bands in which chlorophyll presence can change the spectral response of water were the input data and both bands with original surface reflectance and the ones smoothed with the mean filter were used. Then, the spectral bands were combined in three different ways, enabling the creation of six sets of input data.

The first three sets (I, II and III) are combinations of original spectral bands, while sets IV, V and VI constitute arrangements among the smoothed spectral bands (Table II).

The first combination used seven bands in which water has spectral response (bands 1 to 7) and it was used to evaluate the contribution of the reflectance values throughout the spectrum to the spatial modeling of chlorophyll. In the second type of combination, the aim was to use the bands whose wavelengths are strongly influenced by the presence of chlorophyll, presenting features of scattering (bands of green and red edge), or presenting absorption features, such as the red light region (combinations II and V). The third set aimed to select the regions with the two highest peaks of

reflectance, which have chlorophyll contribution to the define their magnitude, *i.e.*, bands 3 (green) and 6 (red edge), generating the combinations III and VI. The combinations of bands specified are summarized in Table II.

**TABLE II**  
**Input data combination to the JavaNNS.**

Set	Description
I	All original bands (1, 2, 3, 4, 5, 6, 7)
II	Original spectral bands 3, 5, 6
III	Original spectral bands 3, 6
IV	All smoothed bands (1, 2, 3, 4, 5, 6, 7)
V	Smoothed bands 3, 5, 6
VI	Smoothed bands 3, 6

As the purpose was to infer chlorophyll *a*, the output data refer to discrete point values of the pigment concentration in the water body. The image with chlorophyll *a* values was produced in IDRISI Andes, using the geographic parameters of Worldview-2 scene and converting the chlorophyll *a* values from a vector file to the matrix format. Thus, the input data (image) and output (chlorophyll concentration) of the neural network are spatially corresponding.

After the selection of input and output data of the network, the next step was to define the training and validation sets. As stated by Zell et al. (1995), the generalization ability of neural network is one of its greatest advantages, and for better generalization, the authors recommend that the database is divided into three parts:

- Training set: used for learning the neural network. The error of the database is minimized during training;
- Validation set: used to estimate the performance of neural network on the set of patterns that were not trained during learning.
- Test set: used to check the overall performance of the network, which is actually the data set that the trained network will be applied.

Since the test set is the whole scene, it was necessary to define the training and validation sets. They were selected from the survey of continuous flow of the chlorophyll *a*, and consisted of a set of 282 elements for training and 260 for validation of ANN.

To use the JavaNNS, the data must be converted to a file format that the simulator recognizes (\*.pat). Thus, it was necessary to create such files after defining the training and validation samples. For this, a computer program was implemented, that permits to create the patterns files from the data required for ANN training and validation and, from there, it was possible to establish the architecture of the neural network.

By definition, the number of units of the input layer corresponds to the number of input layers. The output layer is the number of expected responses that, in the case of inference, it is just one. For each combination of input data, different network architectures were evaluated, which were constituted of one or two hidden layers and a variable number of processing units in these layers.

The actual information processing within the units in JavaNNS is modeled with an activation function and output function. First, the activation function calculates the unit net inflow from the weighting of the output values of previous units. Then it calculates the new activation of this unit, using the selected activation function, and the output function takes this result to create output value, which will serve as input for the next layer of units (Zell et al. 1995).

It is possible to edit the activation functions of each layer in JavaNNS, allowing greater control in defining the network topology. For this work, following the guidelines of Osório (2006), the units of the input layer have an activation function of the type identity, in the hidden layer activation function was sigmoid (`act_logistic`), and the output layer the activation function was defined as '`act_identityplusbias`', in which there is

a component that adds a gain to the entry value of the function, providing real value of the inferred variable. The author cites this setting as suitable for tasks of approximation/regression, as is the case of this research.

The output functions of units in all layers were defined as identity function, to keep the value processed by the activation function as a result of processing in that unit. This configuration of the units in each layer was maintained for all architectures tested.

The neural network training is the way the net learns that a specific entry corresponds to a given value of output. Therefore, it is through the behavior of a neural network during the training that its suitability for the categorization of available data can be assessed. This means that in addition to network architecture (defined by the number of hidden layers and number of units in each layer) must be chosen the parameters related to the functions of learning, the rules for updating the weights network boot and the criteria to stop the training (Galo 2000).

The training algorithm used was the Resilient Backpropagation (RPROP), which is a variation of the classic backpropagation and has been frequently used for studies of geographical inference (Moreira 2001, Nunes et al. 2007). The RPROP is an algorithm of local adaptation, which requires the definition of three parameters: the initial update value ( $\Delta_0$ ), a maximum update limit ( $\Delta_{\max}$ ) and the decay exponent of the weight ( $\alpha$ ) (Zell et al. 1995).

The first one, as the name implies, is an initial value, from which the weights will begin to be adjusted. The maximum update limit is to prevent the weights to be changed abruptly and the decay exponent of the weight aims to reduce weight, resulting in better generalization (Moreira 2001).

The training algorithm and its parameters were defined based on Zell et al. (1995), Moreira (2001) and Nunes et al. (2007). It was used the RPROP learning algorithm,  $\Delta_0$  equals to 0.1,  $\Delta_{\max}$  equals 40

and the value of  $\alpha$  equals to 4. Some tests were made using  $\Delta_{\max} = 50$  and  $\alpha = 5$ , but it was verified that the change in  $\Delta_{\max}$  did not cause significant changes in the results, while changing  $\alpha$  considerably damaged the representations generated.

Another important component of training is to establish an ending criterion of the iterations of the learning algorithm. Since the simulator allows to operate simultaneously with two sets of standards, training and validation sets, the most appropriate way to check the generalization ability of the network is to monitor the fall of the mean square error (MSE) of the validation set during the training phase, and end the process when it reaches a minimum value (Zell et al. 1995). For each data set, the validation MSE obtained the minimum value in different iterations.

After the training of each network, these were applied to the entire area of the experiment. The JavaNNS provides a result in ASCII format with a specific encoding, so it was necessary to convert it. After that, the text file containing the coordinates and the inferred values were imported into IDRISI Andes and later converted to an image representation, in the same way was made to generate the chlorophyll image collected with fluorometer.

Finally, a visual analysis and statistical analysis of the basic attributes of the data were done, trying to compare them with data collected with the fluorometer to determine the degree of similarity with real data.

## RESULTS

### CHLOROPHYLL CONCENTRATION IN THE EXPERIMENTAL AREA

As stated, from a set of sample with values of chlorophyll fluorescence measured by a field fluorometer and chlorophyll concentration determined in the laboratory, it was possible to fit a regression line for the fluorometric measures calibration. The points used in the calibration include the predominant range of fluorescence

measurements, so the regression line used for calibration was modeled according to the real characteristics of the field data, resulting in the equation (1):

$$[Chl-a] = 0.3501 * F \quad (1)$$

Where  $[Chl-a]$  is the concentration of chlorophyll estimated and  $F$  is the value of chlorophyll fluorescence.

The correlation and the coefficient of determination obtained were satisfactory ( $r = 0.8915$  and  $R^2 = 0.7925$ , respectively), enabling to update the values of the chlorophyll  $a$  concentration measured with the fluorometer in continuous flow on March 18<sup>th</sup> 2011. The basic statistics obtained from a sample of 728 observations is presented in Table III.

**TABLE III**  
Basic statistics of fluorescence obtained by sampling in continuous flow.

	18 <sup>th</sup> March 2011
Minimum	1.57
Maximum	5.22
Mean	2.36
Standard deviation	0.41

It is noticed low values of chlorophyll  $a$  concentration (with an average of 2.36 mg/L) and little variation in these readings ( $s = 0.41$ ). The small variability in the data increases the difficulty of the spatial inference and modeling the phenomenon.

### SPECTRAL REFLECTANCE OF CHLOROPHYLL OBTAINED BY THE IMAGE

The statistical indicators of the Cascata reservoir with and without atmospheric correction in seven spectral bands used in the study are presented in Table IV.

Comparing the digital numbers and reflectance averages in each band, it is realized they exhibit the same variation, reducing their values from band 1 to band 2, then increasing in band 3, decreasing



**TABLE IV**  
**Basic statistics of Cascata's reservoir spectral attributes.**

Spectral Bands	Digital Number		Surface reflectance(*10 <sup>4</sup> )	
	Mean	Std. Deviation	Mean	Std. Deviation
B1 – Coastal	334.6565	9.6281	394.5470	66.2251
B2 – Blue	247.6795	10.3307	388.8598	58.0216
B3 – Green	264.3918	14.7204	428.0365	47.8276
B4 – Yellow	183.5864	16.8027	375.7113	55.6964
B5 – Red	105.9276	14.5465	283.2701	62.5362
B6 – Red Edge	125.9864	28.2601	305.7339	92.5249
B7 – Near Infra Red 1	91.4104	35.1614	204.2784	103.1716

again in the fourth band and so on. The standard deviation did not follow the same behavior, which may indicate with the minimization of atmospheric effects, the combination of different optically active components in the water body causes distinct changes throughout the electromagnetic spectrum.

Figure 2 shows the reflectance image of the spectral band 3 before and after smoothing.

The high variation in the values recorded in the original band 3 of Worldview-2 image (Figure 2a) was smoothed by a mean filter (Figure 2b) and median filter (Figure 2c). However, last filter caused



**Figure 2** - Original reflectance image of band 3 (a), the one smoothed with mean filter (b), and smoothed with median filter (c).

a “wavy” appearance in the scene whereas the mean filter produced a more homogeneous image. So, the images smoothed with mean filter were used in the spatial inference of chlorophyll *a*, together with the original spectral bands.

#### CHLOROPHYLL *a* MAPPING AND ANN PERFORMANCE

The quality of inference depends on appropriated selection of the input data, the training and the validation areas. As shown in Table III, the actual values of chlorophyll have a small variation, so

the areas chosen for training and validation must maintain this small variability. The statistical attributes of the areas of training and validation are presented in Table V.

The statistics presented in Table V indicate that, for both the training and validation data, chlorophyll *a* values vary little around the mean with a standard deviation of about 0.3, although the difference between the maximum and minimum value is approximately 2.2 and 1.5 for the training and validation data, respectively. In this sense, even though the extreme values cannot

**TABLE V**  
Descriptive statistics of the training and validation sets.

	Training	Validation
Minimum	1.568	1.603
Maximum	3.781	3.077
Mean	2.613	2.290
Standard deviation	0.312	0.344
Number of elements	282	260

be considered representative of the sample, it was decided to not remove them, since the use of average values could result in a representation with variability even smaller.

The network training and validation were performed in JavaNNS, in which a number of architectures and learning parameters were tested. For the best results, each set of input parameters of the synthesis of each ANN used and the respective numbers of cycles and MSE are presented in Table VI.

The parameters of the learning algorithm were kept the same in all networks and lower validation errors were obtained in different cycles for each ANN tested. In general, the number of cycles required was relatively small, because it is considered that the training paradigm used consists of an algorithm with faster conversion than the backpropagation one (Braga et al. 2000), with the exception of the second ANN (II - *original bands 3, 5 e 6*) that had 3400 cycles.

The evaluation of the results was done primarily through the indices provided by the simulator, which indicated the quality of the architecture chosen based on the ANN training and validation data. In Table VI it can be noticed that the smoothed data provided validation errors smaller than the original data. Based on MSE, the best configuration was the one based on spectral bands 3 (green), 5 (red) and 6 (red edge) smoothed (V). In this analysis, the selection of bands for specific features of

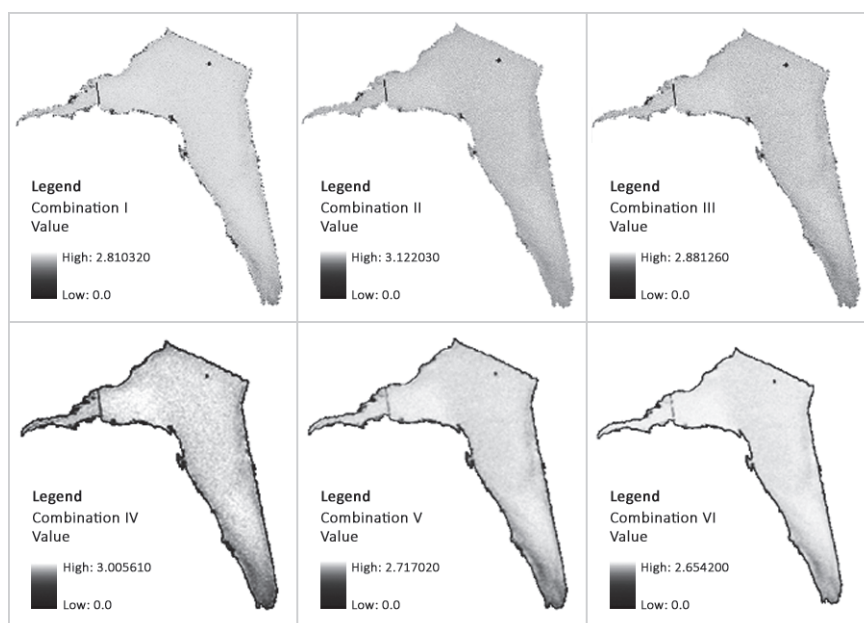
**TABLE VI**  
Descriptive statistics of the training and validation sets.

Combination of input data	Architecture	$\Delta_0$	$\Delta_{\max}$	$\alpha$	# cycles	MSE <sub>validation</sub>
I - <i>all original bands</i>	7-5-4-1	0.1	40	4	400	0.003906
II - <i>original bands 3, 5, 6</i>	3-2-1	0.1	40	4	3400	0.003858
III - <i>original bands 3, 6</i>	2-5-3-1	0.1	40	4	200	0.003780
IV - <i>all smoothed bands</i>	7-5-4-1	0.1	40	4	200	0.002801
V - <i>smoothed bands 3, 5, 6</i>	3-2-1	0.1	40	4	100	0.002733
VI - <i>smoothed bands 3, 6</i>	2-5-3-1	0.1	40	4	100	0.002892

chlorophyll *a* did not provide results far superior to the use of all bands. Thus, it is considered that the main contribution to improve the results of the network is associated with the need for multispectral image processing.

After analyzing the performance of each ANN itself, spatial representations of chlorophyll *a* were generated, also aiming to compare the results obtained with different configurations (Figure 3).

As the validation errors indicated, the networks trained with smoothed data showed results of spatial distribution of chlorophyll *a* visually superior to the networks trained with original data. It is noticed again the importance of image preprocessing, and also how the characteristic of the input data remains in the ANN output, since the grainy appearance of the original images persisted.



**Figure 3** - Spatial representation of values inferred by different configurations of ANN.

The resulting representations with the original bands did not allow observing regions with well-defined chlorophyll concentration differences, while these differences were detected in smoothed images, especially the representations with spectral bands 3, 5, 6 (V) and 3, 6 (VI). In these two results it is clear that the region with lower levels of chlorophyll occur near of the reservoir source, in its south region.

The representations with smoothed images showed well-defined edges with small amounts of chlorophyll and were inferred chlorophyll values

even for a place with no water response, because of a little bridge in the small region in west. In the representations with the original data the bridge region is preserved without values inferred. Thus, it is clear that is required caution in the analysis in edge regions of smoothed images.

The mainly descriptive statistics of the attributes estimated by ANN different architectures (minimum and maximum values, mean and standard deviation) and the values obtained in the field by fluorometer are presented in Table VII.

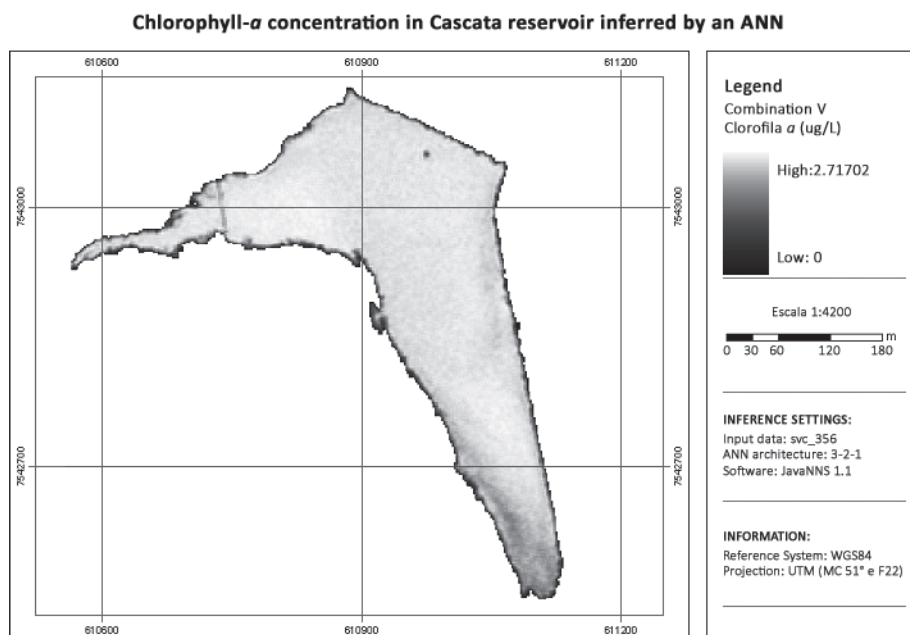
**TABLE VII**  
Descriptive statistics of the attributes estimated by ANN different architectures.

	<i>All original bands</i>	<i>Original bands 3, 5, 6</i>	<i>Original band 3, 6</i>	<i>All smoothed bands</i>	<i>Smoothed bands 3, 5, 6</i>	<i>Smoothed bands 3, 6</i>	Fluorometer
Minimum	0.831	0.624	0.024	0.205	0.254	0.100	1.57
Maximum	2.810	3.122	2.881	3.006	2.717	2.654	5.22
Mean	2.514	2.536	2.519	2.406	2.378	2.380	2.36
Standard deviation	0.268	0.136	0.189	0.570	0.434	0.466	0.41

Table VII shows that none of the networks gave higher values of chlorophyll *a*, concentrating it on the range where there are more data (2 to 3  $\mu\text{g/L}$ ). Again, the representations that closest to the field data were smoothed spectral bands 3, 5, 6 (V) and 3, 6 (VI) in this order. This confirms that from a set of multispectral images processed, the inclusion of bands in regions of the spectrum in which the

spectral response of chlorophyll changes the spectral characteristics of the water provides better results in the inference.

The best performance in all analysis performed was found with the combination of smoothed bands 3 (green), 5 (red) and 6 (red edge) (V). The result of geographical inference of chlorophyll *a* using ANN, for this data set, is presented in Figure 4.



**Figure 4** - Chlorophyll concentration inferred from smoothed spectral bands 3, 5 and 6 of Worldview-2 image.

ANN: Artificial Neural Networks.

## DISCUSSION

The efficiency of a field fluorometry to estimate chlorophyll was certified, providing an adequate amount of data to environmental modeling. However, to obtain reliable models, a critical step is the calibration of the fluorometric data, so the laboratory analysis of water must be carefully executed to enable the fluorometric data represent reality, otherwise it should be clear that it is using a relative measure. In this work, whose water body studied had very small concentrations of the variable in question, it was observed that the techniques for determination of

chlorophyll *a* were appropriate, since the sensing ranges of the fluorometric method (field) and HPLC (laboratory) are superior to those obtained by spectrophotometry.

Regarding the multispectral image, it is necessary to use spectral bands related with the chlorophyll response that was showed by the worst results obtained when all seven spectral bands were used. Also, it is important to reaffirm the importance of atmospheric correction and analysis of the quality of the atmospheric model used. In this case, due the acquisition scene conditions was necessary its smoothing with a mean filter.

For the use of the artificial neural networks is necessary to know the phenomenon to define the activation functions. Moreira (2001) and Nunes (2007) used ANN for inference, but all activation functions of the units were defined as logistics functions, generating an output network between 0 and 1, consistent with the application of these authors, since it sought to infer the degree of aptitude for some activities. In this work, a logistic function would not be appropriate, since it would not provide actual values of chlorophyll *a*.

The definition of training and validation sets should be done carefully, trying to include the main differences of the variable to be inferred, otherwise, the ANN can over-generalize the result. For this work, as the variation of chlorophyll *a* was small, the ability to extrapolate values has not been sufficiently tested.

Finally, it was observed the need for additional analysis to evaluate the performance of the ANN, beyond of validation errors presented by the network. In this research, the use of visual analysis and comparison of statistical attributes were the alternatives found, but for cases where the variations of the model are greater, the correlation analysis can be used. Although this study had a sufficient number of sample elements measured in the field to perform this kind of the analysis, the characteristic of homogeneity of the variable chlorophyll *a* did not allow such operations.

### CONCLUSIONS

The research allowed to evaluate the integrated use of spectral and fluorometric data in spatial inference of chlorophyll *a*. It was noticed that the interaction between fluorometric and multispectral data provided good results when combined in an artificial neural networks structure, allowing to infer the variability of the studied pigment.

Even presenting low and homogeneous values of chlorophyll *a*, it was possible to develop an appropriate model of inference, which represents

the actual variations observed *in situ*. The research allowed the definition of the data set that best fit the reality of the environment studied, which may serve as a basis for studies in other environments.

Finally, it is recommended to evaluate the methodology in areas with significant variations of chlorophyll *a*, aiming to confirm whether the method used for areas with small variations is also effective in cases of high variability.

### ACKNOWLEDGMENTS

We acknowledged the following people and institutions: Prof. Dr. Mauricio Galo and Prof. Dr. Enner Alcântara (FCT/UNESP); Prof. Dr. Edivaldo Domingues Velini and MSc. Tiago Henrique (FCA/UNESP); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Programa de Pós Graduação em Ciências Cartográficas (PPGCC); Fundação de Estudos e Pesquisas Agrícolas e Florestais (FEPAF) and DAEM (Departamento de Água e Esgoto de Marília).

### RESUMO

Considerando a importância do monitoramento de parâmetros da qualidade da água, o sensoriamento remoto é uma alternativa viável à detecção de variáveis limnológicas que possuem propriedades de interação com a radiação eletromagnética, chamadas componentes opticamente ativos (COA). Dentre esses, cita-se a clorofila *a*, que é o pigmento mais representativo da atividade fotossintética em todas as classes de algas. Nesse sentido, o presente trabalho se propôs a desenvolver um método de inferência espacial da concentração de clorofila *a* utilizando Redes Neurais Artificiais (RNA). Para atingir tal objetivo, foi utilizada uma imagem multiespectral e medidas fluorimétricas como dados de entrada. A imagem multiespectral foi tratada, os dados de treinamento e validação da rede foram cuidadosamente selecionados e, a partir disso, foi definida a arquitetura de RNA e parâmetros mais adequados para modelar a variável de interesse. Ao final do treinamento, a rede foi aplicada à imagem e foi



efetuada uma análise de qualidade. Com isso, percebeu-se que a interação entre dados fluorimétricos e multispectrais forneceu bons resultados de aplicação quando combinados em uma estrutura de redes neurais artificiais.

**Palavras-chave:** sensoriamento remoto da água, fluorescência, clorofila *a*, inferência espacial, rede neural artificial.

## REFERENCES

- APHA - AMERICAN PUBLIC HEALTH ASSOCIATION. 1998. Standard methods for the examination of water and wastewater. 20<sup>th</sup> ed., Washington: APHA, 1569 p.
- BARBOSA PCC. 2003. Aplicações de Fluorescência Induzida por Laser em Monitoramento Ambiental. Tese (Doutorado em Física) – Pontifícia Universidade Católica, Rio de Janeiro, 139 f.
- BRAGA AP, LUDEMIR TB AND CARVALHO ABPLF. 2000. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro: LTC, 262 p.
- BUKTON D, O'MONGAIN E AND DANAHER S. 1999. The use of Neural Networks for the estimation of oceanic constituents based on the MERIS instrument. *Int J Remote Sens* 20: 1841-1851.
- CHANG DH AND ISLAM S. 2000. Estimation of Soil Physical Properties Using Remote Sensing and Artificial Neural Network. *Remote Sens Environ* 74: 534-544.
- DANSON FM, ROWLAND CS AND BARET F. 2003. Training a neural network with a canopy reflectance model to estimate crop leaf area index. *Int J Remote Sens* 24: 4891-4905.
- OSÓRIO F. 2006. Dicas de Simulação de Redes Neurais. Disponível em <<http://osorio.wait4.org/oldsite/sadi/dicas/dicas-simulacao-snns.txt>>. Acesso em 10 abr. 2011.
- DIGITAL GLOBE. 2010. World View 2. Disponível em <<http://www.digitalglobe.com/index.php/88/Worldview-2>>. Acesso em 20 de junho de 2010.
- ENNES R. 2008. Potencial das imagens hiperespectrais orbitais na detecção de componentes opticamente ativos no reservatório de Itupararanga. Dissertação (Mestrado em Ciências Cartográficas) UNESP, Presidente Prudente, 126 f. (Unpublished).
- FISCHER I, HENNECKE F, BANNES C AND ZELL A. 2001. Java Neural Network Simulator: User Manual, Version 1.1. Tübingen: University of Tübingen, 33 p.
- GALO MLT. 2000. Aplicação de redes neurais artificiais e sensoriamento remoto na caracterização ambiental do parque estadual morro do diabo. Tese (Doutorado em Ciências da Engenharia Ambiental) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 197 p.
- GROSS L, THIRIA S, CURIE M, FROUIN R AND MITCHELL BG. 2000. Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *J Geophys Res* 105: 3483-3495.
- HAYKIN S. 1999. Redes Neurais: Princípios e práticas. 2<sup>a</sup> ed., São Paulo: Bookman, 900 p.
- HOLLER FJ, SKOOG DA AND CROUCH SR. 2009. Princípios de análise instrumental. 6<sup>a</sup> ed., Porto Alegre: Bookman, 1056 p.
- ITT - VISUAL INFORMATION SOLUTIONS. 2010. ENVI Help. Boulder: ITT.
- JANG JD, VIAU AA AND ANCTIL F. 2004. Neural network estimation of air temperatures from AVHRR data. *Int J Remote Sens* 25: 4541-4554.
- KEINER LE AND BROWN CW. 1999. Estimating oceanic chlorophyll concentrations with neural networks. *Int J Remote Sens* 20: 189-194.
- KEINER LE AND YAN XH. 1998. A neural network model for estimating sea surface chlorophyll and sediments from Thematic Mapper Imagery. *Remote Sens Environ* 66: 153-165.
- LONDE LR. 2008. Comportamento espectral do fitoplâncton de um reservatório brasileiro eutrofizado – Ibitinga(SP). Tese (Doutorado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 223 f.
- MAS JF AND FLORES JJ. 2008. The application of artificial neural networks to the analysis of remotely sensed data. *Int J Remote Sens* 29: 617-663.
- MOREIRA FRS. 2001. Uso e avaliação de técnicas de integração e análise espacial de dados em pesquisa mineral aplicadas ao planalto de Poços de Caldas. Dissertação (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 166 p. (Unpublished).
- NUNES JOR, SAMIZAVA TM, IMAI NN, GALO MLBT AND KAIDA RH. 2007. Aplicação e avaliação de uma rede neural artificial para escolha de áreas para aterro sanitário. *Formação (Presidente Prudente)* 14: 113-125.
- RICHARDS JA AND JIA X. 2006. Remote Sensing digital image analysis: An Introduction. 4<sup>th</sup> ed., Australia: Springer, 439 p.
- SCHILLER H AND DOERFFER R. 1999. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. *Int J Remote Sens* 20: 1735-1746.
- TURNER DESIGNS. 2004. Cyanobacteria detection in water using *in vivo* fluorometry. Disponível em: <<http://www.turnerdesigns.com/t2/doc/appnotes/algal.html>>. Acesso em: 10 de junho de 2010.
- UPDIKE T AND COMP C. 2010. Radiometric use of Worldview-2 imagery. Technical Note. Digital Globe, 16 p.
- WETZEL RG. 2001. Limnology: Lake and river ecosystems. 3<sup>rd</sup> ed., San Diego: Academic Press, 1006 p.
- ZELL A ETAL. 1995. SNNS: Stuttgart Neural Network Simulator: User Manual, Version 4.1. Stuttgart: University of Stuttgart, 338 p.