Enhancing the Unmixing Algorithm through the Spatial Data Modeling for Limnological Studies

Enner Herenio Alcântara [1]
José Luiz Stech [1]
Evlyn Márcia Leão de Moraes Novo [1]
Cláudio Clemente Faria Barbosa [2]
[1] Remote Sensing Division, Brazilian Institute for Space Research
{enner,stech,evlyn}@dsr.inpe.br
[2] Imaging Processing Division, Brazilian Institute for Space Research
claudio@dpi.inpe.br
São José dos Campos, São Paulo, Brazil

1. Introduction

Conventional water quality monitoring is expensive and time consuming. This is particularly problematic if the water bodies to be examined are large. Conventional techniques also bring about a high probability of undersampling (Hadjimitsis et al. 2006). Conversely, remote sensing is a powerful tool to assess aquatic systems and is particularly useful in remote areas such as the Amazon lakes (Alcântara et al., 2008).

Data collected using this technique can provide a synoptic overview of such large aquatic environments, which could otherwise not be observed at a glance (Dekker et al. 1995). However, remote sensing is not easily applied to aquatic environment monitoring mainly because the mixture of the optically active substances (OAS) in the water. Several approaches have been proposed to cope with this issue such as derivative analysis (Goodin et al. 1993), the continuum removal (Kruse et al. 1993), and spectral mixture analysis (Novo and Shimabukuro, 1994; Oyama et al. 2009).

The two first approaches are more suitable for hyperspectral images, whereas the spectral mixture analysis can be used for both hyperspectral and multispectral images. The Spectral Mixture Model (SMM) has largely been used for spectral mixture analysis, uncoupling the reflectance of each image pixel (Tyler et al. 2006) into the proportion of each water component contributing to the signal. The result of a spectral mixture analysis is a set of fraction images representing the proportion of each water component per image pixel. This technique has been applied to TM/Landsat images to determine the concentration of suspended particles (Mertes et al. 1993), chlorophyll-a concentration (Novo and Shimabukuro, 1994); as well as to MODIS images, to determine the chlorophyll-a concentration in the Amazon floodplain (Novo et al. 2006), to characterize the composition

of optically complex waters in the Amazon (Rudorff et al. 2007), and to study turbidity distribution in the Amazon floodplain (Alcântara et al. 2008, 2009a).

Remote sensing data have been extensively used to detect and to quantify water quality variables in lakes and reservoirs (Kloiber et al. 2002). One of the most important variables to monitor water quality is turbidity, because it gives information on underwater light availability (Alcântara et al. 2009b). Although turbidity is caused by organic and inorganic particles, one unresolved issue is to distinguish between them using remote sensing (Wetzel, 2001). The Spectral Unmixing Model (SMM) can, however, be useful to analyze the turbidity caused by inorganic particles and by phytoplankton cell scattering.

This chapter book shows how to improve the well know unmixing algorithm using the spatial data modelling concept and also their applicability in limnological studies.

2. Study Area

The Amazon River basin drains an area of approximately 6×10^6 km², which represents 5% of the Earth surface. The Central Amazon has large floodplains covering around 300,900 km² (Hess et al. 2003), including 110,000 km² of the main stem 'Varzeas' (white water river floodplains; Junk, 1997). At high water, the Amazon River flows into the floodplains, and fills both temporary and permanent lakes which might merge to each other. The 'Lago Curuai' floodplain (Figure 1) covers an area varying from 1340 to 2000 km² from low to the high water. This floodplain is located 850 km from the Atlantic Ocean, near Óbidos (Pará, Brazilian Amazon). Formed by 'white' water lakes (characterized by high concentration of suspended sediments), and 'black' water lakes (with a high concentration of dissolved organic matter, and a low concentration of sediments; Moreira-Turcq et al. 2004). These lakes are connected to each other and to the Amazon River. The floodplain also has 'clear' water lakes filled by both rainwater and by river- water drained from 'Terra Firme' (Barbosa, 2005).

The lakes are connected to each other and to the Amazon River. The Curuai floodplain is controlled by the Amazon River flood pulse (Moreira-Turcq et al. 2004) which creates four states (Barbosa, 2005) in the floodplain-river system (Figure 2): (1) rising water level (January - February), (2) high water level (April – June), (3) receding water level (August – October), and (4) low water level (November – December).

The exchange of water between the floodplain and the Amazon River is shown in Figure 2. When the water level is high, there is very little flow, and the surface water circulation is caused predominantly by wind. In the receding state, the exchange of water between the river and the floodplain is inverted, i.e., the water flows from the floodplain to the river. The water level then drops to the low water state, when the exchange of water between the river and the floodplain is at a minimum. According to Barbosa (2005), during the rising water state, the flow from the river to the floodplain starts at a channel located on its Eastern border, and then migrates to small channels located on its Northwestern side.

The rate of inundation is influenced by the floodplain geomorphology, by the density of floodplain channels, and by the ratio of local drainage basin area to lake area. A majority of 93% of the flooded area in the floodplain is between 2-6 m in depth (using a water level reference of 936 cm). The deepest lake is the 'Lago Grande', and the shallowest are the 'Açaí' and Santa Ninha Lakes (Figure 1-c). An area of about 0.04% of the floodplain is below sea level (with the mean altitude of the floodplain at 9 m above sea level (Barbosa, 2005).
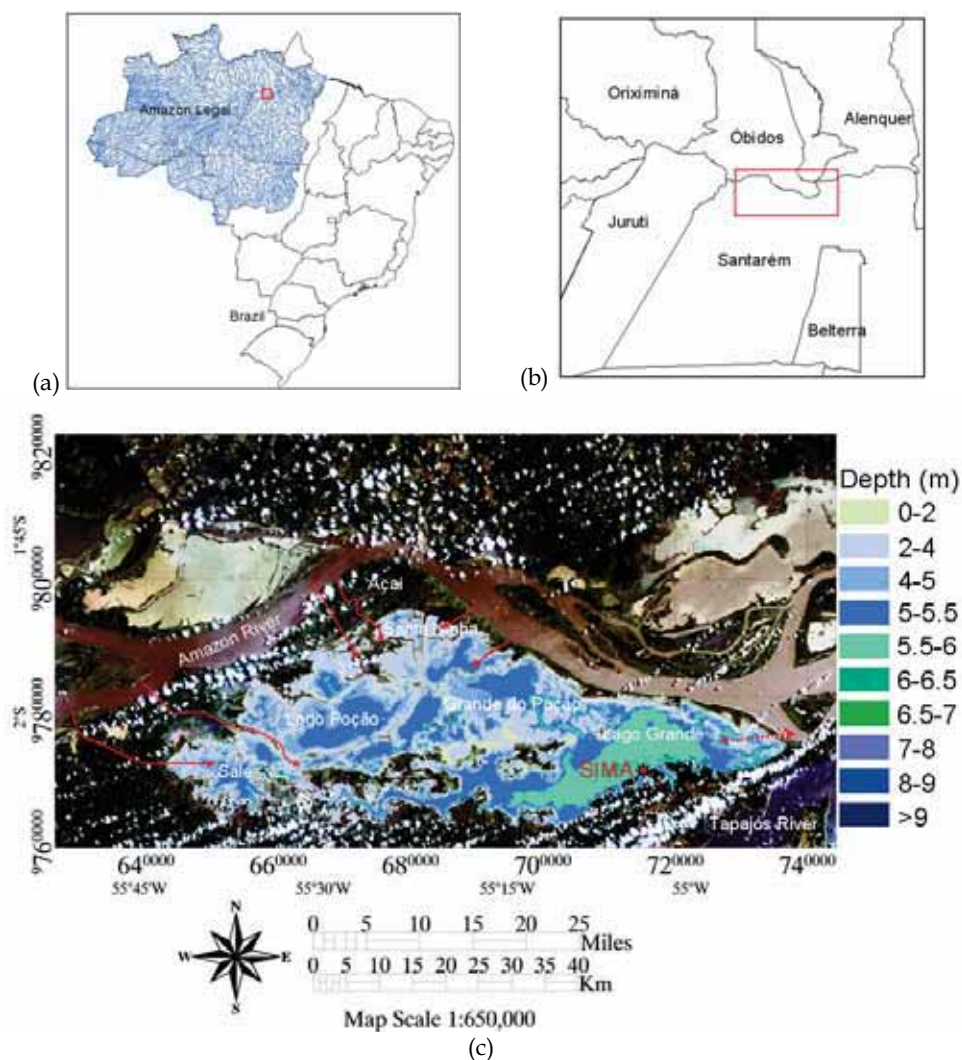
Fig. 1. Legal Amazon location into Brazil (a); (b) Location of Curuai floodplain into Amazon Legal; (c) Landsat-5 Thematic Mapper (normal composition) imagery showing the Curuai Floodplain and their bathymetry. The arrows indicate the main channels of connection Amazon River-floodplain.
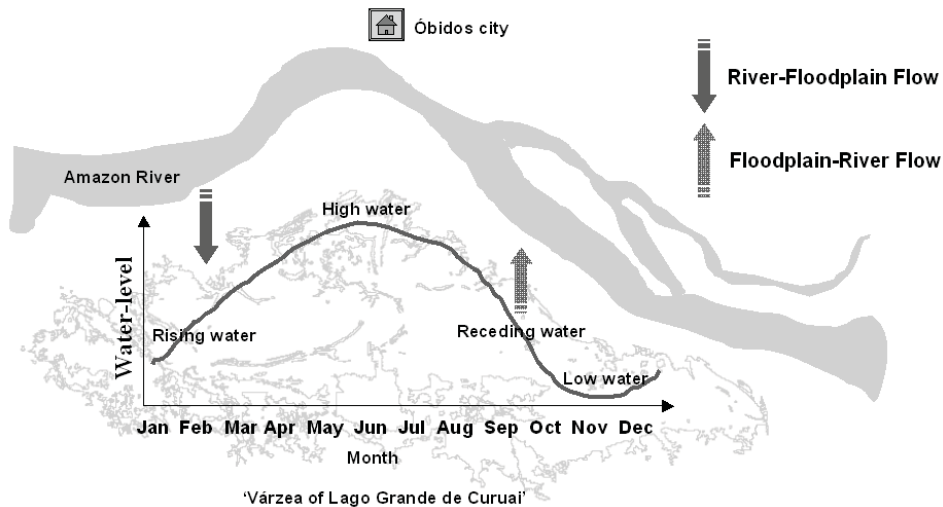
Fig. 2. Theoretical water level dynamic at the Curuai Floodplain (Source: Barbosa, 2005).

3. In-situ Measurements and Remote Sensing Data

The turbidity ground data acquisition was carried out from February 1st to February 14th 2004, during the rising water period. Turbidity measurements were taken at 215 sampling stations using the HORIBA U-10 multi-sensor. This equipment provides turbidity measurements in NTU (Nephelometric Turbidity Unit) with a resolution of 1 NTU. The locations of the sampling stations were determined with the aid of spectral analyses of Landsat/TM images taken at similar water level (Barbosa, 2005). These samples had maximum, minimum and mean values of 569, 101 and 232.29 NTU, respectively.

A Terra/MODIS image, acquired as MOD09 product on February 27th 2004 was used in this study. The spectral bands used in the analyses were band 1 (620-670 nm) and 2 (841-876) with a spatial resolution of 250 m, and bands 3 (459-479 nm) and 4 (545-565 nm) with a spatial resolution of 500 m. The two latter bands were re-sampled to 250 m using MODIS Reprojection Tool software (MRT, 2002).

4. Methodological Approach

The turbidity distribution was assessed using fraction images derived from the Linear Spectral Mixing Model, using four MODIS spectral bands (3 – blue, 4 – green, 1 – red and 2 – near infrared) with a spatial resolution of 250 m. In order to evaluate the turbidity distribution observed in the MODIS fraction images, in-situ measurements acquired during in February 2004 (a few days apart of the MODIS acquisition) were used to apply the Ordinary Least Square (OLS), spatial lag, and spatial error regression models. The kernel estimation algorithm was used to verify the spatial correlation of the in-situ data before performing the regression analyses. A summary of our methodological approach is presented in Figure 3:
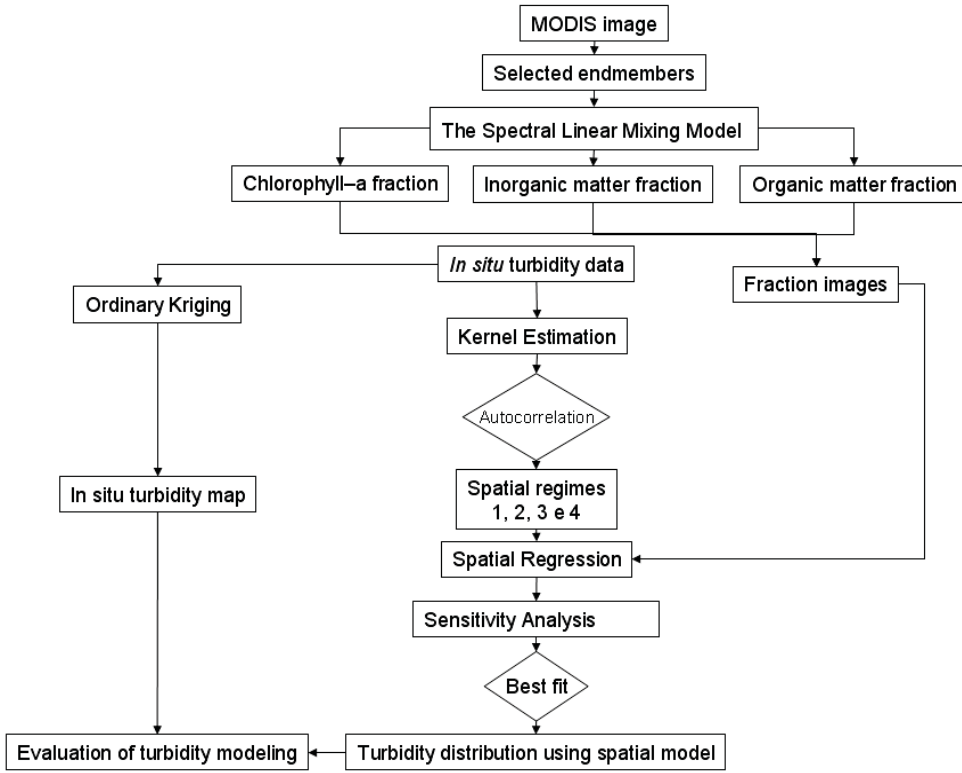
Fig. 3. Flow chart of the methodological approach.

4.1 The SMM

The SMM estimates the proportion of the various surface components present in each image pixel based on their spectral characteristics (Novo and Shimabukuro, 1994). The number of spectral endmembers used in the SMM algorithm must be less than or equal to the number of spectral bands (Tyler et al. 2006). Given these conditions, it is possible to determine the proportion of each component by knowing the spectral responses of pixel components according to equation 1:

$$R_{is} = \sum_{j=1}^{n} f_j r_{ij} + \varepsilon_i \qquad (1)$$

Where: R_{is} is the reflectance at each spectral band i of a pixel with one or more components, r_{ij} is the spectral reflectance of each component j in the spectral band i; f_j is the fraction of the component within the pixel; ε_i is an error term for each spectral band i.

Also, j = 1, 2, ..., n (n = number of components), and i = 1, 2, ..., m (m = number of spectral bands), with the following constraints: $\sum f_j = 1$, and $f_j \geq 0$ for all components.

4.1.1 Endmember selection
As pointed out in (Rudorff et al. 2006), mixtures of dissolved or suspended materials will always occur in natural water bodies. The desired conceptual 'pure' endmembers are hence not accessible for the OAS. Thus, the SMM results will not lead to a complete separation of the fractions. They will rather indicate a relative proportion of each endmember in which the relationship with the actual concentration of a certain OAS will be stronger according to its reflectance spectral dominance.

Some authors have selected the endmembers in a spectral library or laboratory measurements and applied them in the satellite images (Mertes et al. 1993). This approach for selecting the most 'pure pixel' for each component in the water sometimes does not consider the actual characteristics of endmembers found in the local area (Theseira et al. 2003). Thus, some authors collect the endmembers directly in the image, the so called image endmember (Novo and Shimabukuro, 1994).

The endemembers that will be selected in this work are the phytoplankton, inorganic suspended particle and dissolved organic matter laden water. The phytoplankton and the inorganic particle can cause the turbidity in the water and the dissolved organic matter, on the other hand, is a representative manner of non-turbid water. The method of Alcântara et al. (2008) was used to select the endmembers.
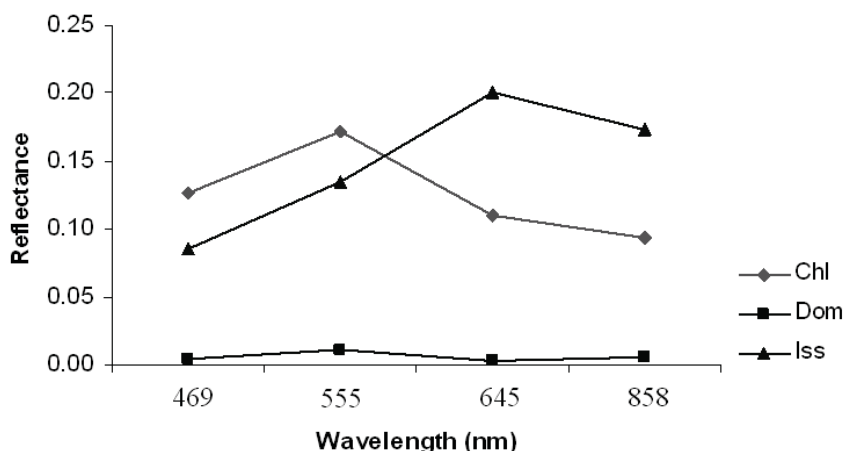

Fig. 4. Image endmembers used as input to run the Linear Spectral Mixing Model on the Terra/MODIS image acquired on 27th February 2004. Chl is the phytoplankton laden water, Dom is the dissolved organic matter laden water, and Iss is the inorganic particle laden water.

Figure 4 shows that the spectral responses of the selected endmembers were quite different. Chl-laden water endmember is characterized by high reflectance in the green band, typical of phytoplankton-laden waters. Dom-laden water endmember presents low reflectance at all wavelengths and the Iss-laden water endmember is characterized by an increasing reflectance towards to visible wavelength, and decreasing slightly at the near infrared.

This model was used to estimate the proportion of each component in the water using the following components: phytoplankton (Chl) laden water, dissolved organic matter (Dom) laden water, and inorganic particle (Iss) laden water. These components (endmembers) were selected based on the dominance of them in the surface water. Then, to estimate the proportion of components into the water we used the following equation:

$$R_{is} = f_{Chl}r_{Chl} + f_{Dom}r_{Dom} + f_{Iss}r_{Iss} + \varepsilon_i \qquad (2)$$

Where, R_{is} is the reflectance at the i_{th} band at sampling station s; f_{Chl}, f_{Dom} and f_{Iss} are the fractions phytoplankton laden water, dissolved organic matter-laden water and inorganic matter-laden water, respectively; r_{Chl}, r_{Dom} and r_{Iss} are the reflectance of phytoplankton laden water, dissolved organic matter-laden water and inorganic matter-laden water, respectively and ε_i is the error at the ith spectral bands.

The main idea of this study was to estimate the proportion of each of the three components in the surface water. Ground data of Chl, Dom, and Iss concentration in water was then used to understand the distribution of turbidity in the Curuai floodplain lake. The results obtained with the SMM were compared with in-situ turbidity measurements through the interpolation, using a geostatistical method, called Ordinary Kriging.

4.2 Ordinary Kriging

A turbidity map was generated with the in-situ data, for subsequent comparison with the fraction images generated using the SMM. Thereby, it could be verified to which extent the two data sets matched. To produce such a reference map, Ordinary Kriging was used, interpolating in-situ turbidity measurements, as described in (Isaaks and Srivastava, 1989).

Ordinary Kriging is a technique of making optimal, unbiased of regionalized variables at unsampled locations using the structural properties of the semivariogram and the initial set of data values. The calculation of the Kriging weights is based upon the estimation of a semivariogram model, as described by equation 3:

$$\gamma(h) = \frac{1}{2}Var[z(s+h) - z(s)] \qquad (3)$$

Where: $\gamma(h)$ is an estimated value of the semivariance for the lag h. The estimation of a semivariogram model relies on two important assumptions: (1) the parameter $\gamma(h)$ exists and is finite for all choices of h and s, and that it do not depend on s. The Ordinary Kriging estimator is presented in:

$$Z(x, y) = \sum_{i=1}^{n} w_i z_i \qquad (4)$$

Where n is the number of considered measures, zi are the corresponding attribute values, and wi are the weights.

The experimental semivariogram was fitted with various theoretical models (spherical, exponential, Gaussian, linear and power) using the weighted least square method. The theoretical model that gives minimum RMSE is chosen for further analysis. In this case, the fitted model was based on Gaussian model (Table 1).

Water level	Anisotropy direction	Structure	Nugget	Sill	>range	<range	Model
Rising	94°	1°	619	114	16436	ε	Gaussian
		2°		7770	17924	16436	
		3°		1480	∞	17924	

Table 1. The semivariogram parameters used to interpolate the in-situ turbidity.

The adjustment on Gaussian model suggests the existence of smooth spatial variance pattern in the of study site (Burrough and Mcdonnell, 1998). The reference map was then used to evaluate the result obtained with the SMM. The equation 5 presents the fitted model used to interpolate the turbidity distribution during the rising water phase.

$$\gamma(h) = 619 + 114[Gau(\sqrt{\left(\frac{h_{94°}}{\varepsilon}\right)^2 + \left(\frac{h_{216°}}{16436}\right)^2})] + 7770[Gau(\sqrt{\left(\frac{h_{94°}}{17924}\right)^2 + \left(\frac{h_{216°}}{16436}\right)^2})] +$$
$$1480[Gau(\sqrt{\left(\frac{h_{94°}}{17924}\right)^2 + \left(\frac{h_{216°}}{\infty}\right)^2})] \qquad (5)$$

Where $\gamma(h)$ is the semivariance at the lag h, $h_{n°}$ is the semivariance due to angles of anisotropy, ε is the range of the lower anisotropy angle, and 'Gau' is the fitted Gaussian fitted model.

4.3 Spatial regression between the SMM and in situ turbidity

In general, OLS models have been used in research (Tyler et al. 2006). This approach, however, does not consider the spatial autocorrelation of samples within an aquatic system. When spatial autocorrelation is not considered in regression analysis, the significance of parameters can be overestimated, and the existence of large-scale variations might lead to spurious associations (Anselin, 1988). In our study, a method called Kernel Estimation (KE) was applied to in-situ turbidity, thereby testing for the existence of spatial autocorrelation (which is indicative of different spatial regimes).

The KE is a common analysis tool to determine the local density of point events and create a field representation. The Kernel is computed, using a Gaussian function, k, which considers m events (i.e. turbidity samples) $\{u_i, ...u_i + m - 1\}$ contained in a bandwidth τ around u and the distance d between the position and the ith sampling (Figure 5), through the functions represented by equation 6:

$$\hat{\lambda}_\tau(u) = \frac{1}{\tau^2} \sum_{i=1}^{n} k\left(\frac{d(u_i, u)}{\tau}\right), d(u_i, u) \leq \tau \tag{6}$$
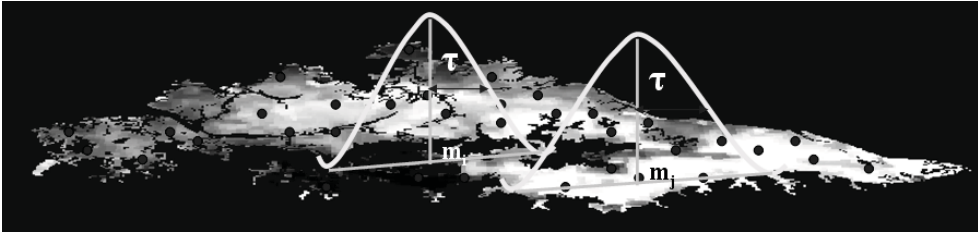

Fig. 5. Kernel estimation of the spatial turbidity pattern.

To improve the kernel estimation we used an adaptive bandwidth version which locally adjusted bandwidth values at different points within the floodplain. Thereby, it ensured that the bandwidth always contained a minimum number of samples, improving estimate precision (Bailey and Gatrell, 1995).

The technique was also applied to the OLS, a model compiled using global in-situ turbidity data to evaluate the spatial regression models. The OLS does not take into account spatial dependence among samples and consequently, different spatial regimes. However, spatial error and lag models were developed for each identified spatial regime, using the KE algorithm, thus including spatial dependence which tends to inflate the variance of OLS regression (Bailey and Gatrell, 1995).

To assess the relationship between the fraction images and in-situ turbidity two approaches were adopted The first, assumes that the variance of the disturbance term is constant; we start with the OLS model:

$$y = X\beta + \varepsilon, \text{ with } \varepsilon \sim N(0, \sigma^2) \tag{7}$$

Where y is an (Nx1) vector of observations on a dependent variable taken at each of N locations, X is an (NxK) matrix of exogenous variables, β is an (Kx1) vector of parameters, and ε is an (Nx1) vector of disturbances.

The second uses two alternative forms of spatial dependence models (Bivand, 1998), the spatial lag model (presented in equation 8) and the spatial error (presented in equation 9).

$$y = \rho W y + X\beta + \varepsilon \tag{8}$$

$$y = X\beta + u, \text{ with } u = \lambda Wu + \varepsilon \tag{9}$$

Where λ is a scalar spatial error parameter, and u is a spatially autocorrelated disturbance vector. These two models can also be related to each other in the Common Factor model.

According to (Bivand, 1998), however the spatial lag and spatial error models can only be combined for estimation if the neighborhood specifications (here: the W matrices of the lag and error components) differ. The spatial lag model is clearly related to a distributed lag interpretation, in that the lagged dependent varible, Wy, can be seen as equivalent to the sum of a power series of lagged independent variables stepping out across the map, with the impact of spillovers declining with successively higher powers of ρ (Bivand, 1998).

Conversely, the spatial error model is a special case of a regression with a non-spherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence (Baltaqi, 2003).

Spatial lag and spatial error dependence tests allowed determining which of the two models was more suitable to the data. Subsequently, spatial regression models were created for each spatial regime. Measurements of Log likelihood fit, Akaike information criteria (AIC) and Schwarz criteria (SC) for both models (OLS and spatial models) to verify if the inclusion of spatial parameters improve the regression model.

The Schwarz criterion addresses the problem of selecting one of a number of models of different dimensions (Bailey and Gatrell, 1995). However, some authors (Burrough and Mcdonnell, 1998; Bailey and Gatrell, 1995) suggest the use of AIC to evaluate the best fit. The Akaike information criterion is expressed in equation 10 (Pan, 2001).

$$AIC = -2 * LIK + 2k \tag{10}$$

Where, LIK is the log of the maximized likelihood and k is the number of regression coefficients. A small AIC value suggests a high suitability of the tested model.

4.4 Evaluation of turbidity estimations

Having selected the best model, the next step was to apply it to the Terra/MODIS image to estimate the turbidity distribution. The resulting map was then used as a reference to evaluate the model. Random positions were selected on the image to run correlation analyses so as to assess the potential spatial model and the RSME. At each random geographical position, a $3x3$ pixel window was averaged, obtaining both the measured and modeled value and computing the correlation.

5. Results and Discussion

Figure 6-a shows that the water, of the entire Curuai floodplain lake was rich in inorganic matter (Iss), with a particularly high proportion in the Poção lake (Figure 6-b). The images illustrating the distribution of dissolved organic matter (Dom) revealed that this was particularly apparent in the Salé lake, and in the littoral region (i.e. the region between water and forest). This is mainly due to the fact that some organic matter in decomposition is transported into the floodplain by the water during the rising , phase (Figure 6-c). The

phytoplankton (Chl) fraction represented a high proportion of the 'Grande de Curuai' and 'Grande do Poção' lakes (Figure 6-d). As described in (Barbosa, 2005), the water coming from the Amazon River was more enriched with phytoplankton than that in the Curuai floodplain; the proportion of the water rich in Chl was found to increase from the East to the West.
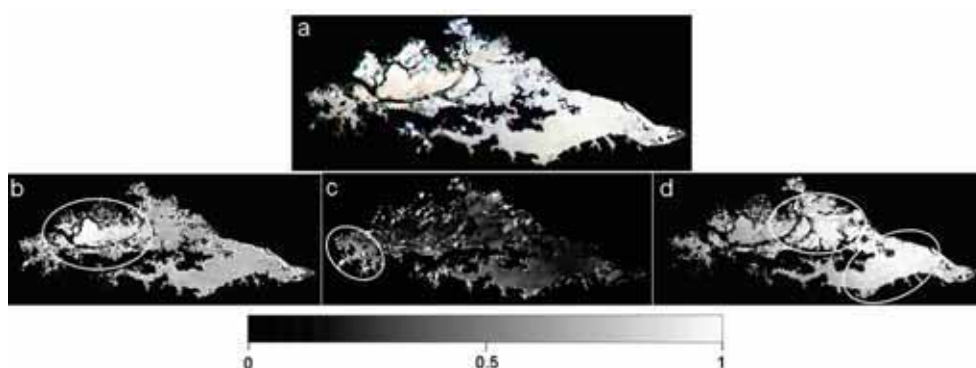

Fig. 6. Terra/MODIS imagery of the Curuai floodplain (a), and proportion images for the tested endmembers: Iss (b), Dom (c) and Chl (d) compiled with the SMM algorithm.

The fraction images (Iss in the red channel, Dom in the blue channel, and Chl in the green channel) unfolded a qualitative picture of the turbidity distribution within the Curuai floodplain lake (Figure 7-a). Figure 7-b shows the SMM error image of the Region 1 (Figure 7-a) which represents the most turbid water within the floodplain. This regions has an area of extremely high turbidity (see the circle in Figure 7-b) which actually is due to cloud cover. Region 2 (Figure 7-a) evidence of the mixture of water masses with dominant proportions of Chl and Iss, and low turbidity (Figure 7-b). Region 3 (Figure 7-a), finally, was dominated by water with a high proportion of Chl, and moderate turbidity. A mixture of water masses containing high concentrations of Chl and Iss lead to a low error of the SMM. Conversely, a high amount of Chl entering the floodplain through the main channel from the Amazon River leads to a moderate error in the unmixing model. Due to the occurrence of a transition zone between the aquatic and the terrestrial environments, higher errors occurred at the edges.

However, these results are qualitative. To obtain suitable quantitative results, an OLS regression model was applied to the SMM results and the in situ turbidity data. Subsequently, we checked for any signs of spatial autocorrelation between samples to prevent the problem of spurious associations, using the result obtained by the kernel estimator algorithm.

The results of the OLS regression model, using all 215 turbidity samples collected, and the fractional abundance of the OAS are poorly correlated ($R^2 = 0.10$, $p < 0.05$). This is presumable due to the fact that the floodplain received water from different sources (rain, black water and white water) when the water level rose. The mixture of water masses caused a high heterogeneity in the spectral response of the surface water, causing high standard deviation in turbidity measurements, what highlights the importance of including the autocorrelation factor in the regression analysis.
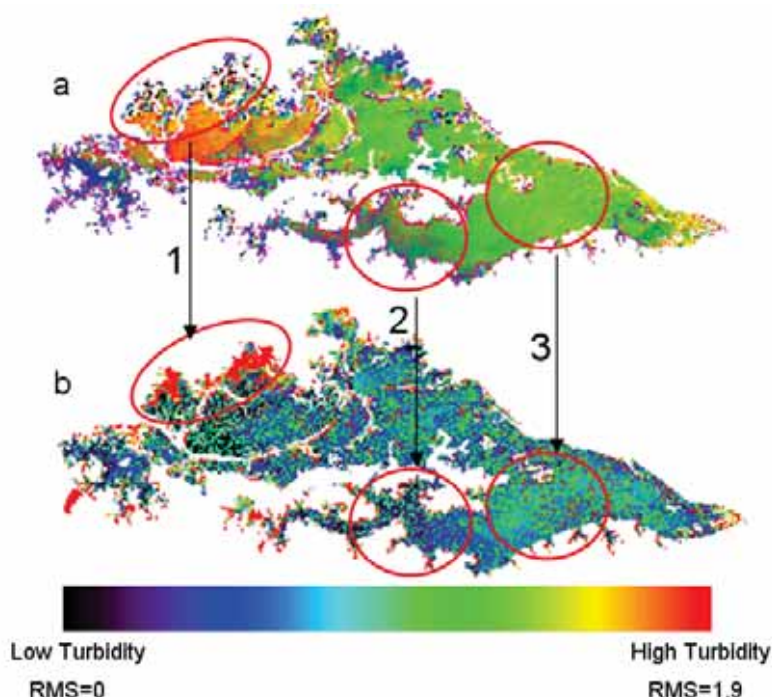
Fig. 7. (a) The composition of the image fractions derived from the SMM (Iss in red channel, Chl in green and Dom in blue), in comparision with the the RMS image (b).

If spatial dependence was verified the OLS method would lose validity (Anselin, 1988). In accordance with Bailey and Gatrell (1995) the use of the non-spatial regression models (OLS), may or may not be suitable because it assumes a stationary water conditions in the period of ground sampling. Also, the OLS model does not consider the presence of spatial autocorrelation among samples distributed within the floodplain. Hence, in order not to overestimate the significance of parameters, autocorrelation must be considered. Likewise, large-scale variations may induce spurious associations.

To check the spatial autocorrelation between turbidity samples, we applied the Kernel estimator, and then separated the samples of fractional abundance of the OAS and in-situ turbidity data in spatial regimes. These cluster the turbidity data from the whole floodplain by their spatial dependence (Bailey and Gatrell, 1995).

The KE revealed that there were four spatial regimes of the Curuai floodplain lake turbidity at the rising water level (Figure 8-a). The kernel also shows four regions of density from high (region 1) to low density (region 4). The region 1 was characterized by the highest spatial dependence in the study area. In this area, the spatial regime is particularly abundant 1(Figure 8-b) and includes the largest number of samples.

Region 2 included a mixture of spatial regimes 1 and 2, and represented the second largest spatial dependence. Region 3 was characterized by regimes 3 and 4, and region 4 included a mixture of all regimes (Figure 8-b). The number of samples grouped in each spatial regime was 64, 54, 51 and 45 for regimes 1, 2, 3, and 4 respectively.
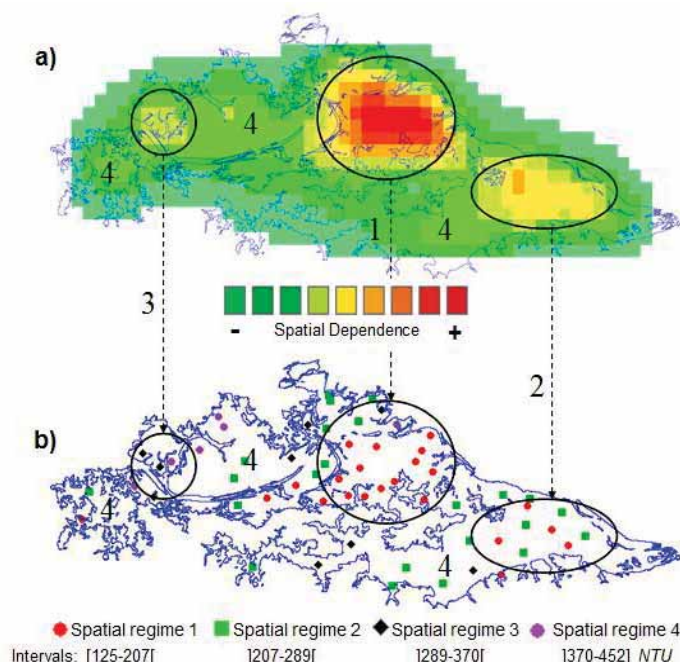
Fig. 8. (a) Spatial dependence of in-situ turbidity samples, as defined by Kernel algorithm, (b) spatial regimes occurring within the Curuai floodplain lakes with corresponding turbidity intervals values (NTU).

The different regimes are brought about by different types of water input when the Amazon River enters into the floodplain. The water arrives in different ways, and at different times. As a result, the water entering the different sections of the floodplain has variable physical and chemical properties.

Table 2 indicates that the application of the spatial regression model increased the R^2 in relation to the OLS model. This is presumably due to the turbidity data being characterized by spatial auto-correlation, as observed in figure 8-a. The OLS model did not take into account the spatial dependence among samples, thereby causing an overestimated significance of the parameters.

	OLS	Spatial regression							
		Spatial Lag Model				Spatial Error Model			
	Global	Re1	Re2	Re3	Re4	Re1	Re2	Re3	Re4
R2	0.10	0.31	0.27	0.61	0.71	0.32	0.30	0.95	0.79
LIK	-295.79	-95.07	-101.88	-35.97	-22.78	-94.93	-101.69	-35.50	-20.70
AIC	599.58	200.15	213.76	81.94	55.57	197.86	211.38	79	49.41
SC	607.54	205.61	219.44	82.93	54.53	202.22	215.92	79.79	48.57

Table 2. Results of the OLS and spatial regression models, with R2 being the coefficient of determination, LIK being the log of the maximized likelihood, AIC being the Akaike information criterion, and SC being the Schwarz criterion.

The approach to estimate the turbidity data in spatial regimes shows that in all regimes the best fit was obtained using the spatial error model. According to Anselin (1988) the fit on the spatial error model suggests that the spatial effect is a noise derived from the interplay of several variables not included in the model.

The spatial regime 1 (Re1) has the poorest correlation compared to the others ($R^2 = 0.10$, p < 0.001), with LIK, AIC and SC being -295.79, 599.58, and 607.54 respectively. Conversely, the spatial regime 3 (Re3) shows the best results with LIK = -35.50, AIC = 79 and SC = 79.79, and R2 = 0.95, p < 0.001, using the spatial error model (Table 2). The equations of the four spatial regimes (Re1, Re2, Re3, and Re4) are listed below in the equations 11, 12, 13, and 14 respectively:

$$Turbidity = 99.72 + 221.71 Iss + 199.46 Chl + 186.53 Dom \qquad (11)$$
$$Turbidity = 311.61 - 254.14 Iss - 23.87 Chl + 22.98 Dom \qquad (12)$$
$$Turbidity = 327.19 + 42.62 Iss - 246.85 Chl - 136.06 Dom \qquad (13)$$
$$Turbidity = 428.86 - 154.66 Iss - 231.23 Chl - 94.05 Dom \qquad (14)$$

Where, Iss is the inorganic-laden water fraction, Chl is the phytoplankton laden water fraction and Dom is the dissolved organic matter-laden water fraction.

Equation 13 indicates the influence of the OAS proportion on this method to quantify turbidity. According to equation 13, the turbidity in this region will be lower when the water has high proportion of phytoplankton than inorganic particles. Conversely, in areas where those types of water are particularly abundant, the modeled turbidity will be smaller. Where the Iss fraction dominates, turbidity will be high. This suggests that the model will perform better when turbidity is determined mainly by suspended inorganic particles.

Before applying the equation 13 to the MODIS image fractions, a sensitivity analysis was carried out using Pearson's correlation analysis for each spatial regime (Figure 9).

This sensitivity analysis revealed statistically significant correlations $R^2 = 0.65$ (p < 0.05, n = 20) and $R^2 = 0.40$ (p < 0.05, n = 20) for the spatial regimes 3 [Figure 9 (c)] and 4 [Figure 9 (d)] respectively. In contrast, analyses for the spatial regimes 1 [Figure 9 (a)] and 2 [Figure 9 (b)] were not significant with $R^2 = 0.005$ (p < 0.05, n = 20) and $R^2 = 0.004$ (p < 0.05, n = 20) respectively (Figure 9). The best fit was hence found in the spatial regime 3.
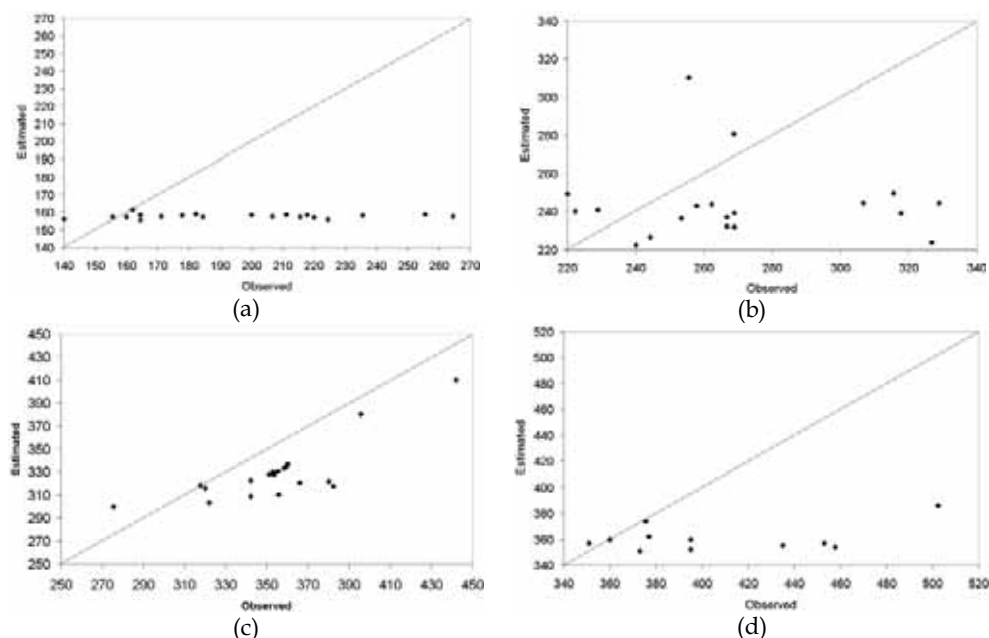
Fig. 9. Correlation between turbidity estimated using the spatial model and observed turbidity: (a) spatial regime 1; (b) spatial regime 2; (c) spatial regime 3; (d) spatial regime 4.

The main reason for this poor statistical performance is probably related to the instantaneous image acquisition and the sampling turbidity design. Both cause a difference in turbidity conditions, which makes a comparison of the image and the in-situ data difficult. In addition, the nonlinearity presented in the relationship between the loads of phytoplankton, dissolved organic matter, and inorganic matter in the water made the unmixing processes difficult.

Figure 10 shows the turbidity distribution resulting from the application of this model to the entire floodplain (Figure 10-a) as well the in-situ turbidity distribution (Figure 10-b). The spatial distribution of the turbidity is similar, especially in regions 1, 2, and 3.

Region 1 is characterized by the highest turbidity compared to the other regions, and by small depths when the water level rises (Barbosa, 2005), and is the main pathway from the Amazon River into the floodplain. The water then flows through many channels, located in the Northwest region of the floodplain, with sufficient energy to keep the inorganic particles in suspension. Winds promote the sediment re-suspension and increment turbidity here.

Region 2 is also characterized by small depths but is protected from wind. As a result, turbidity is related to depth rather than wind perturbation (Carper and Bachmann, 1984; Booth et al. 2000). In Region 3, turbidity varied between 203 and 305 NTU, and the spatial model was able to estimate turbidity values within the range measured in-situ. According to Barbosa (2005), this is one of the regions with the highest depths in the Curuai floodplain lake. It is not subjected to intense wind perturbation due to this characteristic. Instead, suspended particles settle into the bottom of the lake, decreasing the possibility of re-suspension.

Figure 10-c indicates that the spatial model overestimated turbidity in regions 4 and 6, and underestimated it in region 5. The overestimation in 4 can be explained by cloud cover contamination. Region 5 is in a transition zone (natural barrier) that separates the floodplain lake in two larger zones, (i) zone Northwest and (ii) zone Southeast (Barbosa, 2005). This transition zone causes a turbidity underestimation due to mixture of different water masses. The overestimation in region 6 can be related to a large mixture caused by a water inlet owing to a rising water level. The water inlet is particularly pronounced in region 6 (Barbosa, 2005).
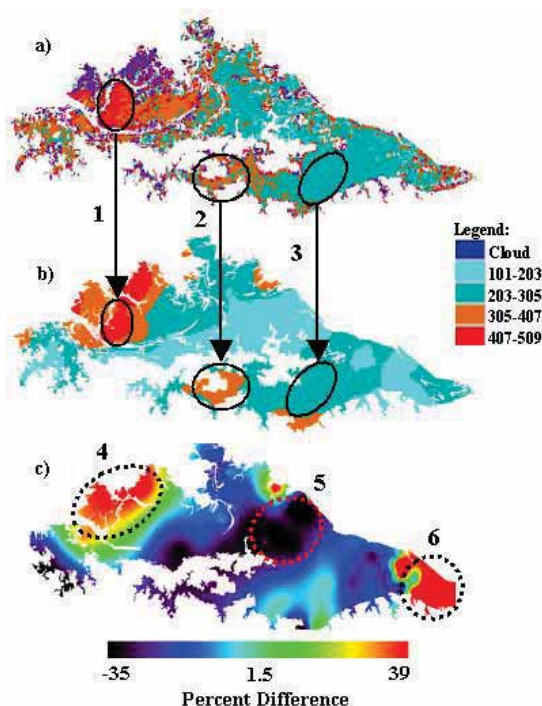

Fig. 10. The turbidity distribution (NTU) in the Curuai floodplain, (a) using the spatial regression model, using the OK algorithm (b) and (c) is in percent difference between (a) and (b).

The samples to evaluate the spatial model were collected the entire lake, except for regions 4 (due to cloud cover), 5 (natural barrier), and 6 (overestimation due to intense water flowing through this channel), as pointed out in Figure 7. The evaluation of the model resulted in a value of $R^2 = 0.90$ (p = 0.05; n = 60) and RSM of 17 NTU for the turbidity model under the spatial regime 3 (Figure 11). Previous studies of turbidity using the SMM in high water level (Alcântara et al., 2008) show that the modeled turbidity had a correlation $R^2 = 0.62$ (p > 0.005; n = 20).

This difference in performance can be explained by the presence of different water types in the floodplain. Their presence makes it difficult to apply one model, adapted to a given regime for the entire region. Another possible error source is the in-situ turbidity-sampling

scheme. Due to the size of the lake, it was not possible to get all the samples on a single day. Instead, the data was collected during a 13-day period, in which local turbidity may have been affected by changes in wind intensity, light field, and other environmental factors affecting the lake hydrodynamics. The highest errors were encountered in more turbid waters, as opposed to clear waters, presumably due to a high mixture of the OAS present in the floodplain.
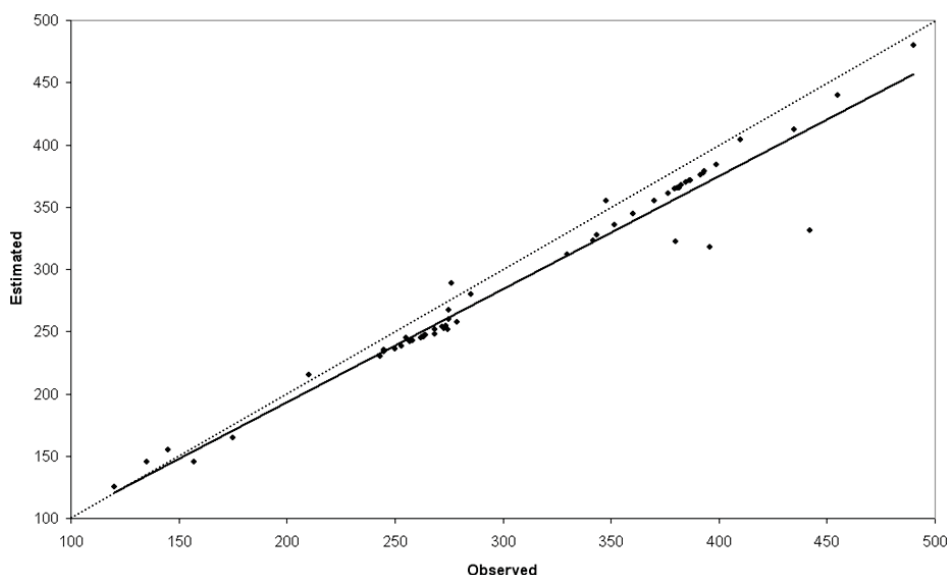

Fig. 11. Evaluation of the turbidity estimated through the spatial model (equation 13) and in-situ turbidity (NTU), interpolated using the OK algorithm.

Since the MODIS image was instantaneously acquired on 27 February 2007, the modeled turbidity does not account for environmental changes, which can affect in-situ conditions. In spite of this drawback, the unmixing model showed a good potential to assess the turbidity in continental aquatic systems. This potential could be improved using hyperspectral remote sensing imagery (Rudorff et al. 2007), since features in spectral responses could then be detected in more detail (Rudorff et al., 2006; Fraser, 1998; Brando and Dekker, 2003).

6. Conclusions

This present work evaluates the suitability of the spectral unmixing model to map the turbidity distribution in the Curuai floodplain. The main conclusions are:
1. The fraction images for the endmembers selected directly from the MODIS image based on dominance of water components allowed assessing the turbidity in the Curuai floodplain lake.
2. Owing to non-linearity in the Amazon floodplain waters, the unmixing model does not work in an optimal way. This is also due to autocorrelation presented in the study area.

 3. It was clear that the presence of autocorrelation in limnological studies that use spatial distributed samples and this paper shows the possible way to solve the problem of autocorrelation between samples spatially distributed.

4. The spatial regression between the results obtained from MODIS fraction imaging with the map generated from in situ data using the Ordinary Kriging approach seems to be useful to estimate water turbidity.

5. The modeling of autocorrelation helps to improve the applicability of the SMM to map the turbidity distribution in high complexity water bodies, such as the Amazon floodplain.

7. Acknowledgements

8. References

Alcântara, E.H.; Barbosa, C.; Stech, J.; Novo, E.; Shimabukuro, Y. (2009a). Improving the spectral unmixing algorithm to map water turbidity distributions. Environmental Modelling & Software 24, 1051-1061.

Alcântara, E.H.; Novo, E.; Stech, J.; Lorenzzetti, J.; Barbosa, C.; Assireu, A.; Souza, A. (2009b). The turbidity behavior in an Amazon floodplain. Hydrol. Earth Syst. Sci. Discuss., 6, 3947–3992.

Alcântara, E.H.; Stech, J.L.; Novo, E.M.L.M.; Shimabukuro, Y.E.; Barbosa, C.C.F. (2008). Turbidity in the Amazon floodplain assessed through a spatial regression model applied to fraction images derived from MODIS/Terra. IEEE T. Geosci. Remote 46 (10), 2895-2905.

Anselin, L. (1988). Spatial Econometrics: Methods and Models; Kluwer Academic: Dordrecht, Netherlands, pp. 284.

Barbosa, C.C.F. (2005). Sensoriamento remoto da dinâmica de circulação da água do sistema planície de Curuai/ Rio Amazonas. (PhD. Thesis); INPE: São José dos Campos, Brazil, pp. 255.

Boardman, J. W., Kruse, F. A., Green, R. O. (1995). Mapping target signatures via partial unmixing of AVIRIS data: in Summaries, Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1 (1), 23-26.

Baltaqi, B.H. (2003). A companion to theoretical econometrics. Blackwell: Texas, USA, pp. 709.

Bailey, T.C.; Gatrell, A.C. (1995). Interactive spatial data analysis; Prentice Hall: Upper Saddle River, USA, pp. 432.

Bivand, R. (1998). A review of spatial statistics techniques for location studies. In Proceedings of the CEPR symposium on New Issues in Trade and Location, Lund, Sweden, 1-21.

Booth J.G.; Miller R.L.; Mckee B.A.; Leathers R.A. (2000). Wind-induced bottom sediment resuspension in a microtidal coastal environment. Cont. Shelf Res. 20 (7) 785-806.

Brando, V.E.; Dekker, A.G. (2003). Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. IEEE T. Geosci. Remote. 41 (6), 1378-1387.

Burrough, P.A.; Mcdonnell, R.A. (1998). Principles of geographical information systems; Oxford University Press: New York, USA, pp. 333.

Carper, G.L.; Bachmann, R.W. (1984). Wind resuspension of sediments in a Prairie Lake. Can. J. Fish. Aquat. Sci 41, 1763-1767.

Dekker, A.G.; Malthus, T.J.; Hoogenboom, H.J. (1995). The remote sensing of inland water quality. In Advances in Environmental Remote Sensing; Danson, F.M., Plummer, S.E., Eds., Wiley, pp. 123-142.

Fraser, R.N. (1998). Hyperspectral remote sensing of turbidity and chlorophyll a among Nebraska Sand Hills lakes. Int. J. Remote Sens. 19 (8), 1579-1589.

Goodin, D.G.; Han, L.; Fraser, R.N.; Rundquist, D.C.; Stebbins, W.A.; Schalles, J.F. (1993). Analysis of suspended solids in water using remotely sensed high resolution derivative spectra. Photogramm. Eng. Rem. S. 59 (4), 505-510.

Hadjimitsis, D.G.; Hadjimitsis, M.G.; Clayton, C.; Clarke, B. (2006). Determination of Turbidity in Kourris Dam in Cyprus Utilizing Landsat TM Remotely Sensed Data. Wat. Res. Manag. 20 (3), 449-465.

Hess, L.L.; Melack, J.M.; Novo, E.M.L.M.; Barbosa, C.C.F.; Gastill, M. (2003). Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. Remote Sens. Environ. 87 (4), 404-428.

Isaaks, E.H.; Srivastava, M.R. (1989). An introduction to applied geostatistics; Oxford University Press: New York, USA, pp. 561.

Junk, W.J. (1997). The Central Amazon Floodplain: ecology of a pulsing system; Springer Verlag: Berlin, Germany, pp. 525.

Kloiber, S.M.; Brezonik, P.L.; Olmanson, L.G.; Bauer, M.E. (2002). A procedure for regional lake water clarity assessment using Landsat multispectral data. Remote Sens. Environ. 82 (1), 38-47.

Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sens. Environ. 44 (2-3), 145-163.

Mertes, L.A.K.; Smith, M.O.; Adams, J.B. (1993). Estimating suspended sediment concentrations in surface waters of the Amazon River wetlands from Landsat images. Remote Sens. Environ. 43 (3), 281-301.

Moreira-Turcq, P.; Jouanneau, J.M.; Turcq, B.; Seyler, P.; Weber, O.; Guyot, J.L. (2004). Carbon sedimentation at Lago Grande de Curuai, a floodplain lake in the low Amazon region: insights into sedimentation rates. Palaeo. Palaeoc. Palaeoe. 214 (1-2), 27-40.

MRT. (2002). MODIS Reprojection Tool User's Manual; Department of Mathematics and Computer Science South Dakota School of Mines and Technology. USGS EROS Data Center, pp. 1-57.

Novo, E.M.L.M.; Shimabukuro, Y.E. (1994). Spectral mixture of inland tropical waters. Int. J. Remote Sens. 15 (6), 1351-1356.

Novo, E.M.L.M.; Barbosa, C.C.; Freitas, R.M.; Shimabukuro, Y.E.; Melack, J.M.; Pereira Filho, W. (2006). Seasonal changes in chlorophyll distributions in Amazon floodplain lakes derived from MODIS images. Limnology 7 (3), 153-161.

Oyama, Y.; Matsushita, B.; Fukushima, T.; Matsushige, K.; Imai, A. (2009). Application of spectral decomposition algorithm for mapping water quality in a turbid lake (Lake Kasumigaura, Japan) from Landsat TM data. ISPRS Journal of Photogrammetry and Remote Sensing 64 (1), 73-85.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. Biometrics 57 (1), 120-125.

Rudorff, C.M; Novo E.M.L.M.; Galvão L.S; Pereira-Filho W. (2007). Derivative analysis of hyperespectral data measured at field and orbital level to characterize the composition of optically complex waters in the Amazon (in Portuguese). Acta Amaz. 37 (2), 269-280.

Rudorff, C.M.; Novo, E.M.L.M.; Galvão, L.S. (2006). Spectral mixture analysis for water quality assessment over the Amazon floodplain using Hyperion/EO-1 Images. Revista Ambi-Água 1 (2), 65-79.

Theseira, M.A.; Thomas, G.; Taylor, J.C.; Gemmell, F.; Varjo, J. (2003). Sensitivity of mixture modelling to end-member selection. Int. J. Remote Sens. 24 (7), 1559-1575.

Tyler, A.N.; Svab, E.; Preston, T.; Présing, M.; Kovács, W.A. (2006). Remote sensing of the water quality of shallow lakes: A mixture modelling approach to quantifying phytoplankton in water characterized by high-suspended sediment. Int. J. Remote Sens. 27 (7-8), 1521- 1537.

Wetzel, R.G. (2001). Limnology: lake and river ecosystems, 3rd Ed.; Academic Press: San Diego, USA, pp. 1006.