Relationship between water transparency and physical-chemical variables in lakes of the western Amazon, Brazil

Giselle Vanessa Trevisan[1*] and Dayson José Jardim-Lima[2]

[1]Centro de Ciências do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais, Av. dos Astronautas, 1758, 12227-010, São José dos Campos, São Paulo, Brazil. [2]Programa de Pós-graduação em Geografia Física,Universidade de São Paulo, São Paulo, São Paulo, Brazil. *Author for correspondence. E-mail: givtrevisan@yahoo.com.br

ABSTRACT. Water transparency is one of the main indicators of seasonal changes from the water level in lacustrine systems, but other environmental spatial and temporal variables can act jointly on water transparency. The diagnosis of interactions among physical-chemical variables and water transparency can be made by fitting statistical models based on multiple regression analysis. The objective of this study is to evaluate the set of limnologic variables that best predict the seasonal variation in water transparency in lakes of the western Amazon. These data were collected in both the drought and flood seasons in 78 lakes of the Mamirauá Reserve for Sustainable Development (RDSM). The delineation of the sampling followed the choice of variables, the analysis of premises imposed by the linear regression model and the validation of the model. The variables that best fitted the model were: water level, bottom temperature, conductivity, and pH for both seasons analyzed (R^2 = 0.64). The resulting model suggests that the physical-chemical variables influence the re-suspension process and sedimentation of particulate material together with the seasonal variation of the water level at the lake system of RDSM.

Keywords: water transparency, regression analysis, lakes, Amazon, Mamirauá Reserve for Sustainable Development.

Variações entre transparência de água e variáveis físico-químicas em lagos da Amazônia Ocidental, Brasil

RESUMO. A transparência da água é um dos principais indicadores das flutuações sazonais do nível d'água em sistemas lacustres, porém outras variáveis ambientais espaço-temporais podem, conjuntamente, atuar sobre a transparência da água. O diagnóstico das interações entre variáveis físico-químicas e a transparência da água pode ser delineado pelo ajuste de modelos estatísticos fundamentados em análise de regressão múltipla. O presente estudo teve como objetivo avaliar o conjunto de variáveis limnológicas que melhor prediz a variação sazonal da transparência da água em lagos na Amazônia Ocidental. Os dados foram coletados durante os períodos de seca e cheia em 78 lagos da Reserva de Desenvolvimento Sustentável Mamirauá (RDSM). O delineamento seguiu a escolha das variáveis, análise das premissas impostas pelo modelo de regressão linear e validação do modelo. As variáveis que melhor se ajustaram ao modelo foram: nível da água, temperatura de fundo, condutividade e pH para ambos os períodos de análise (R^2 = 0,64). O modelo resultante sugere que o processo de ressuspensão e sedimentação do material particulado influenciam as variáveis físicas e químicas juntamente à variação sazonal no nível da água nos sistemas de lagos da RDSM.

Palavras-chave: transparência da água, análise de regressão, lagos, Amazônia, Reserva de Desenvolvimento Sustentável Mamirauá.

Introduction

The Amazon floodplain is one of the most extensive wetlands on Earth. Its area has an extension of 300,000 km^2, considering only the Central Amazon section (IRION et al., 1997). It is a flat region, with a height gradient of around 120 m along its entire extension. Due to this low gradient, water level variations in the main channel between 7 and 13 m correspond to an inundation area of 20 to 100 km perpendicular to the river's margins (COSTA, 2005).

The rivers and floodplains of the Amazon basin constitute a complex of channels, rivers, islands and depressions that are permanently modified by sedimentation and transport of suspended material through whitewater rivers (MERTES et al., 1996). Most of the lakes in these systems are formed by the lateral displacement of the main channel of meandering rivers (SIOLI, 1984) and their chemical composition depends on the entry of water masses originated from the associated river, from the tributaries of *terra firma*

and from rainfall (FORSBERG et al., 1988; JUNK, 1997).

The physical, chemical and biological processes occurring in the lakes vary according to season (flood and drought). During the flood season, surface waters from the rivers enter the lakes through channels remaining in the floodplains, and during the drought this stored water is released naturally, influencing the geochemistry of adjacent rivers (ALCÂNTARA et al., 2008). Generally, during the flood period the suspended material is deposited, increasing water transparency, and during the dry season the vertical mixture caused by winds re-suspends the sediment material, which turns the water more turbid (SCHMIDT, 1973).

According to Wetzel (2001), turbidity (the opposite of water transparency) is a visual property of water, constituting a reduction or absence of light in the water column due to suspended particles, usually inorganic, originated from soil erosion in the river basin and from sediment re-suspension in the bottom. In Amazon lakes, turbidity usually varies with the flooding pulse (ALCÂNTARA et al., 2008) and, for the same time of year, the depth of water transparency can vary significantly, reaching a variation of 0.25 m to 1.3 m in drought season to 0.75 m to 2.2 m in flood season (QUEIROZ, 2007). This behavior suggests the existence of other environmental variables that, in conjunction with the flooding pulse, act on the change in water transparency in Amazon lakes.

The study of the dynamics of water transparency in natural lakes is important because of its consequences for the aquatic biota (RUSSEL et al., 2001). A high concentration of suspended sediment stimulates the occurrence of floating and emerged plants, which reduces the concentration of submersed plants (CARVALHO et al., 2005). The increase of this variable could also cause changes in the bio-dynamic cycles, reducing the phytoplankton community as it interferes in the speed and intensity of photosynthetic activity and influencing negatively on the predation capacity of the ichthyofauna due to its interference in visibility (BUKATA et al., 1995; RADKE; GAUPISCH, 2005).

Aiming to delineate a model that would express the relationship between water transparency in lakes of the Mamirauá Reserve for Sustainable Development, Amazonas State, and limnologic variables which would manifest the dynamics of lacustrine systems, a set of statistical methods was applied to determine the significance of parameters corresponding to the selected variables by multiple regression analysis.

Material and methods

Study area

The Mamirauá Reserve for Sustainable Development (RDSM, Figure 1) is located between the geographical coordinates S 2°48' to 2°54' and W 64°53' to 65°03', in the confluence of the Solimões and Japurá rivers, and present the characteristic várzea mosaic of water bodies and forest types (QUEIROZ, 2007). With a total area of 1,124,000 ha, it is the largest exclusive protection reserve of the Amazon floodplain.
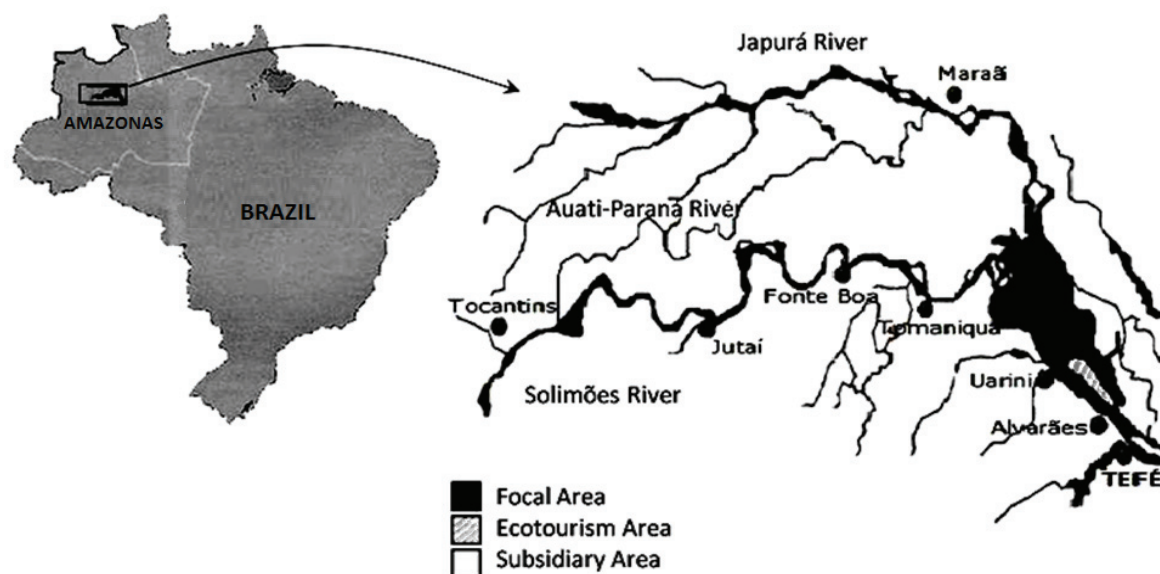
Figure 1. Localization of the Mamirauá Reserve for Sustainable Development, Amazonas State.

Two types of floodplain are found at the RDSM: (1) areas located between the Auati-Paraná and Paraná-Aranapu rivers (Subsidiary Area), which make up 85% of the entire reserve, constituted mainly by terrains of Pleistocene origin, and (2) the floodplains between Paraná do Aranapu and the confluence of the Solimões and Japurá rivers (Focal Area), of Holocene origin (AYRES, 1995). Within the RDSM focal area there are around 830 lakes, distributed among nine lake systems: Mamirauá, Jarauá, Ingá, Liberdade, Tijuaca, Horizonte, Boa União, Aranapu and Barroso.

Preliminary analysis indicates that the lake systems Mamirauá and Jarauá are those with most limnologic information available (QUEIROZ, 2005, 2007). For that reason, both systems were selected for our study.

The Mamirauá system, located at the entry of the Focal Area, is connected during the drought only with the Japurá river and the Jarauá system, which is located at the center of the focal area, connected with rivers Solimões and Japurá during the entire hydrologic year. According to Queiroz (1999), the Jarauá includes approximately 60% of the lakes the entire RDSM focal area.

Physical and chemical parameters

Data on the precise location of the lakes at RDSM and the limnologic parameters were acquired from the database of the Mamirauá Institute for Sustainable Development (IDSM), collected during studies of fish communities at RDSM (QUEIROZ, 2007) in 78 lakes from lake systems Mamirauá and Jarauá (Table 1).

Table 1. Medium values of physical-chemical variables in 78 lakes of RDSM used in this study.

No	SEC	NL	TS	TF	CD	pH	OD	FCA
1	1.2	11.2	27.7	27.7	86.8	6.6	0,0	0.3
2	0.8	12.0	27.2	27.2	94.5	6.8	0,0	0.4
3	2.4	12.5	27.4	27.4	89.9	6.7	0,3	0.6
4	2.0	12.5	27.2	27.0	95.4	6.8	0,2	0.3
5	2.6	14.2	27.9	27.9	92.3	6.8	0,4	0.4
6	1.5	13.6	27.8	27.8	91.0	6.5	0,0	0.2
7	1.6	13.6	27.6	27.5	89.0	6.5	0,0	0.3
8	1.2	13.6	27.4	27.6	96.8	6.7	0,0	0.3
9	1.9	13.6	27.7	26.5	83.0	6.4	0,3	0.0
10	2.0	12.0	27.9	27.6	95.6	6.7	0,2	0.4
11	2.0	11.6	28.1	26.8	79.4	6.4	0.2	0.0
12	2.0	11.6	28.1	26.8	79.4	6.4	0.2	0.0
13	1.9	9.2	27.7	26.9	92.0	6.7	0.0	0.0
14	0.7	9.1	28.0	27.9	95.0	6.7	0.3	0.0
15	1.8	8.9	28.1	27.2	110.0	6.8	0.1	0.0
16	1.1	11.2	28.1	27.1	90.0	6.5	0.2	0.0
17	1.1	8.8	29.4	27.5	111.0	7.1	0.6	0.0
18	2.0	13.6	27.6	26.7	82.2	6.6	0.2	0.0
19	2.4	13.7	27.8	26.8	94.0	6.8	0.1	0.0
20	1.3	13.7	27.9	27.2	95.5	6.8	0.6	0.0
21	1.2	13.7	27.9	27.4	95.5	6.8	0.4	0.0
22	1.9	9.3	29.0	29.0	89.0	6.4	0.2	0.0
23	2.6	8.6	28.3	27.5	94.0	6.5	0.0	0.0
24	0.7	11.5	32.0	28.3	86.3	6.9	0.5	0.0
25	1.3	11.6	28.2	26.3	83.4	6.6	0.0	0.0
26	0.9	11.7	27.7	27.0	89.7	6.8	0.2	0.0
27	0.9	11.7	27.7	27.1	89.5	6.8	0.2	0.0
28	1.3	11.7	27.6	27.2	93.1	6.6	0.2	0.0
29	2.5	11.8	28.0	26.4	81.7	6.7	0.0	0.0
30	1.4	11.8	28.4	27.4	92.8	6.7	0.3	0.0
31	0.6	11.8	29.2	28.9	75.4	6.9	1.0	0.1
32	0.7	12.0	32.0	28.6	87.6	6.8	0.0	0.0
33	0.4	12.0	27.7	27.1	93.1	7.0	0.5	0.1
34	1.3	12.1	27.9	26.7	93.4	6.5	0.0	0.1
35	2.5	13.7	27.9	28.7	95.5	6.8	0.4	0.0
36	1.8	13.7	28.1	27.8	91.3	6.7	0.3	0.0
37	2.5	13.8	28.4	27.5	92.4	6.8	0.4	0.0
38	0.2	5.9	32.3	31.3	53.9	6.9	4.5	0.0
39	0.3	5.5	34.0	31.0	88.7	6.9	4.4	0.0
40	0.2	8.0	31.3	31.3	63.3	6.9	3.5	0.0
41	0.2	7.0	29.6	29.5	93.2	7.2	4.5	0.1
42	0.2	7.5	30.5	31.9	91.6	7.3	3.8	0.0
43	0.2	8.9	30.7	31.5	92.7	6.8	2.3	0.0
44	0.3	6.0	32.8	30.7	85.3	6.8	4.7	0.0
45	0.2	5.5	35.1	31.4	115.6	6.6	5.4	0.0
46	0.2	8.0	32.0	30.9	71.3	6.7	3.5	0.0
47	0.7	7.9	30.2	29.4	77.6	6.5	1.1	0.0
48	0.5	7.9	29.2	29.2	70.6	6.7	4.5	0.0
49	0.3	12.0	30.5	28.7	97.6	6.8	3.9	0.0
50	0.8	7.2	32.8	31.5	75.8	6.7	3.3	0.0
51	0.5	7.5	30.2	30.9	89.6	7.2	6.1	0.0
52	0.9	8.9	31.2	30.6	93.5	6.7	3.0	0.0
53	0.8	7.1	35.5	30.1	92.3	7.0	5.6	0.0
54	0.5	5.4	31.3	31.0	79.7	6.4	3.7	0.0

Continue…

…continuation

No	SEC	NL	TS	TF	CD	pH	OD	FCA
55	0.8	5.4	31.8	30.8	191.5	6.6	2.4	0.0
56	0.4	5.4	31.5	30.0	143.4	7.0	3.2	0.0
57	0.5	7.6	32.4	29.8	87.3	6.9	3.3	0.0
58	0.5	7.6	32.4	29.8	87.3	6.9	3.3	0.0
59	0.4	5.5	31.3	31.3	82.9	6.5	3.7	0.0
60	1.0	7.9	26.9	26.9	83.0	6.2	0.0	0.0
61	0.7	7.7	29.9	29.7	110.6	6.7	1.1	0.0
62	0.7	7.5	31.8	28.3	68.8	6.5	1.5	0.0
63	0.9	7.2	31.9	31.3	99.2	7.1	3.1	0.0
64	2.1	13.6	27.6	27.1	80.4	6.4	0.4	0.1
65	2.0	6.4	27.4	26.0	94.0	6.8	0.1	0.0
66	1.7	12.0	27.0	27.0	92.9	6.7	0.0	0.5
67	2.5	11.8	27.5	27.1	80.0	6.7	0.0	0.0
68	2.5	14.2	27.8	28.9	94.2	6.8	0.1	0.0
69	1.0	12.1	27.5	29.0	94.1	6.9	0.7	0.0
70	1.9	8.8	27.9	27.6	85.5	6.5	0.2	0.0
71	1.3	11.8	27.4	27.4	82.6	6.8	0.2	0.1
72	1.6	12.0	26.9	26.8	92.3	6.6	0.0	0.9
73	1.3	8.9	28.2	26.3	83.4	6.6	0.2	0.0
74	0.2	5.9	32.3	32.3	53.9	6.9	4.5	0.0
75	0.2	7.2	31.3	31.6	75.1	6.9	3.2	0.0
76	0.3	7.1	29.9	29.5	96.6	6.8	2.1	0.0
77	0.9	9.4	31.2	31.1	93.5	6.7	3.0	0.0
78	0.7	8.0	29.6	28.2	111.2	6.7	0.7	0.0

*No: number of lakes; SEC: visual disappearance depth of the Secchi disk (m); NL: water level (m); TS: surface temperature (^{0}C); TF: bottom temperature (^{0}C); CD: water electric conductivity (μs cm^{-1}); pH: hydrogenionic potential; OD: dissolved oxygen (mg L^{-1}); FCA: flow of water current (m s^{-1}). **No: in bold typefaces - lakes of Mamirauá system; not bold - Jarauá system. ***1-47: were collected during the flood season; 48-78 at dry season.

The location of the lakes was made using a GPS instrument, and the water level (NL) was collected daily by a measurement at a river gauge located inside the reserve (November 1992 to March 2001). Water transparency data, measured by the visual disappearance depth of the Secchi disk (SEC), were collected during the dry (NL = 5 m) and flood (NL = 15 m) periods in 78 lakes (n-sampling) of the Mamirauá and Jarauá systems (SIST) within the RDSM. Aside from these parameters, in this study the following data were used: water electric conductivity (CD), surface temperature (TS), bottom temperature (TF), hydrogen-ion potential (pH), dissolved oxygen concentration (OD) and water current flow (FCA).

Model construction

Accounting for 80% of the total n-sampling (training data: 63 samples) were randomly selected and used for the model construction step. The other 20%, accounting for 15 samples, were used in model validation analysis.

The assumptions of any regression model, independence and normality of the dependent variable Y and constant variance of the residues, were investigated respectively by the Kolmogorov-Smirnov (SIEGEL, 1975) and Levene (KUTNER et al., 2004) statistical tests. The Pearson index (KUTNER et al., 2004) was used to measure the correlation between every pair of explaining variables and between each of them and the dependent variable.

The Best Subset and the Forward Stepwise techniques (KUTNER et al., 2004) were used for the selection of variables based on linear regression model

with a single dependent variable. Based on the adjustment criteria R_p^2, R_a^2 and C_p (Best Subset) and FSt (Forward Stepwise), these techniques can define the subset of explaining variables that best predict the values of the dependent variable for new observations.

The values verified by the R_p^2 criterion are based on the existence of a lower proportion of variation in Y for a given X and on the number of variables included in the regression model; hence the maximum value of R_p^2 occurs when all variables are in the model. Because of this, the analysis using the R_p^2 criterion should not be limited to the best value, but rather the value where the addition of variables did not cause an increase in R_p^2. The R_a^2 criterion is analyzed in the same way. The C_p criterion is based on the mean square error and the total variance. The best fit of variables occurs when the lower value of C_p is near the number of parameters in the determined model. Based on F statistics, the Forward Stepwise test analyzes the input and output of many variables and combinations of these until the establishment of a constant variable in the model. A problem linked to this test is the inclusion of highly correlated variables in the model.

The quality of interaction from each variable in the models was evaluated by multiple regression analysis, observing the p-value from each variable and the determination coefficient (R^2) of the estimated models.

Despite being part of the same complex, the Solimões and Jarauá systems have differences in connectivity during the drought season. That encouraged us to test the existence of variations in each system, considering them as a binary qualitative variable. So a complete model weighting with the systems variable and a reduced model without the

systems influence were adjusted. The significance of the coefficients of the adjusted regressions was verified by a linear F statistics test for both the complete and the reduced model.

The validation of the chosen model was done using two procedures: (i) comparison of the intercept coefficients (β_0 - point of intersection of the line with the ordinate), the angular coefficients of regression (β_1, β_2, β_3, … , β_x - measure of variation associated with explanatory variable when the dependent variable is changed in one unit) and the respective standard deviations ($s\{\beta\}$) originated from the training and validation data; (ii) analysis of the β hypothesis of the linear regression between the validation observed values (Y) and the predicted values (Ŷ), calculated from β parameters of the chosen model for the validation data. All the analysis was done using the software STATISTICA 6.0.

Results and discussion

As required when using linear regression analysis, the normality of the dependent variable was analyzed. Water transparency (Y) presented a normal distribution only when log-transformed (Kolmogorov-Smirnof, $D_{calculated} = 0.110 < D_{critical} = 0.153$). The result of this change also increased the correlation coefficient between Y and the explaining variables (Table 2).

Table 2. Pearson's Correlation coefficients of variable dependent (Y and LogY) and explaining variables.

	Y	LogY	NL	TS	TF	CD	pH	OD	FCA
Y	1.00								
LogY	0.94	1.00							
NL	0.65	0.68	1.00						
TS	-0.67	-0.70	-0.72	1.00					
TF	-0.69	-0.77	-0.69	0.82	1.00				
CD	0.03	0.09	-0.09	0.02	-0.02	1.00			
pH	-0.37	-0.45	-0.19	0.32	0.39	0.09	1.00		
OD	-0.71	-0.79	-0.73	0.83	0.85	-0.07	0.44	1.00	
FCA	0.28	0.27	0.36	-0.38	-0.28	0.03	-0.08	-0.29	1.00

*Y: water transparency; LogY: variable log-transformed.

The independent variables best related with Y also presented a high correlation among them, meaning that they are not suitable for being jointly used in the model (Table 2). The adjustments among explaining variables and variable Y agree with literature findings (ALMEIDA; MELO, 2009; CRISTOFOR et al., 1994; FREIRE et al., 2009; MELO; SOUZA, 2009), except for the flow. The relationship between flow and water transparency should be inverse, due to the increment of suspended sediments in the water column caused by the swirl (ALMEIDA; MELO, 2009; CRISTOFOR et al., 1994; MELO; SOUZA, 2009). However, the increased turbidity in Amazon lakes occurs during the low water period, when the process of

sedimentation is interrupted by the same re-suspension induced by wind (ALCÂNTARA et al. 2010; BOURGOIN et al., 2007). In this study, 60% of data were collected in the flood season (Table 1), which could explain the positive relationship between flow and water transparency.

The statistical tests analyzed presented three sets of different variables for the adjustment of the model, where criteria R_p^2 and R_a^2 obtained the same selection of variables. The best models generated for criteria R_p^2, R_a^2, C_p and FSt are presented on Table 3, and the acceptance and/or rejection of selected variables of these models, determined by the p-value, are listed at Table 4.

Table 3. Significant values for the selection of variables (p < 0.05) according to criteria R_p^2, R_a^2, C_p and FSt.

Criterion	E	COT	TS	TF	CD	pH	COD	FCA
R_p^2	5	0.248		-0.277	0.130	-0.192	-0.266	
R_a^2	5	0.248		-0.277	0.130	-0.192	-0.266	
C_p	4	0.335		-0.419	0.160	-0.246		
FSt★	2			5.905			93.284	

"E" represents the number of effects/steps needed for the selection.

Table 4. Multiple regression parameters (p-value and R^2) between water transparency and selected variables on the models generated by R_p^2, R_a^2, C_p and FSt criteria.

	Model 1 (R_p^2 e R_a^2)		Model 2 (C_p)		Model 3 (FSt)	
	p	R^2	p	R^2	p	R^2
COT	0.067	0.68	0.009	0.64	0.018	0.64
TF	0.078		0.002			
CD	0.104		0.043			
Ph	0.034		0.004			
COD	0.107				0.002	

Model 1, defined by criteria R_p^2 and R_a^2, reached the best value of the determination coefficient (R^2) that is, the greater measure of adequacy of fit of the regression with five variables included in the model:

$$Y = 3,188 + 0,030^*NL - 0,056^*TF + 0,002^*CD - 0,312^*pH - 0,05\star OD \qquad (1)$$

Nevertheless, one observes that only variable pH presented a significant acceptance probability ($p_{pH} \leq 0.05$). This demonstrates that the high value of R^2 obtained in Model 1 represents the cumulative effect of a certain number of highly correlated variables (Table 1 and 4), which induced the rejection of this model.

With only two included variables, the Model 3:

$$Y = 2,164 - 0,073^*TF - 0,089^*OD \qquad (2)$$

Obtained the value of a determination coefficient R^2 equivalent to Model 2, validated by criterion C_p with four variables:

$$Y = 4,36 + 0,041\star NL - 0,085\star TF + 0,003\star CD - 0,398\star pH \qquad (3)$$

The adequacy of a smaller number of predictor variables that explain the total variation in the dependent variable Y in Model 3 would be advantageous were it not for the fact that FSt statistics may adjust models that include highly correlated variables. In this case, the only two variables included in Model 3 (TF and OD) have the highest correlation coefficient (Pearson = 0.85), which determined the exclusion of this model.

The high correlation between the TF and OD variables led to the impossibility of inserting them in the models due to redundancy of information, which probably caused the reduction of the individual significance of variables selected in Model 1. Therefore, the adjustment criterion C_p used along with the Best Subset technique generated the most appropriate model (Model 2), with a good coefficient of determination and a significant inclusion of all selected variables (Table 4).

Model 2 predicts that 64% of water transparency variation in RDSM lakes is explained by a set of four variables: positively by the variation in water level and conductivity – a higher NL and CD, a greater transparency of water; and negatively by the variation in bottom temperature and pH – a lower TF and more acidic lake water, a greater value of depth of the visual disappearance of Secchi disk. This does not mean that the relationship found between the independent variables is standard for other lake systems. In RDSM lakes, some studies showed an inverse relationship between these variables. Chaves et al. (2005) found that during the flood in the *floodplains* lakes of RDSM the waters are cooler and more alkaline than in the dry season. According to Henderson et al. (1999), during the flooding when the water level reaches its maximum, the conductivity in the lakes of Jarauá and Mamirauá systems is lower ($\mu = 88$) than during the dry season ($\mu = 94$).

The complete and reduced models adjusted to the binary qualitative variables Solimões and Jarauá were generated from Model 2, as follows:

Complete model:

$$Y = \beta_0 + \beta_1 NL + \beta_2 TF + \beta_3 CD\beta + \beta_4 pH + \beta_5 SIST + \beta_7 TF.SIST + \beta_8 CD.SIST + \beta_9 pH.SIST \tag{4}$$

Reduced model:

$$Y = \beta_0 + \beta_1 NL + \beta_2 TF + \beta_3 CD + \beta_4 pH \tag{5}$$

According to the F test used to determine the existence of variation between the Solimões and Jarauá systems, there is no difference between the complete and reduced models at the level of

significance $\alpha = 0.05$ ($F^\star = 0.579 \leq F_{critical}$ 2.389). Therefore, the validation process was applied to the reduced model, which corresponds to Model 2.

All analysis used in the validation step demonstrated the good fit of the selected variables of Model 2. The residues of the multiple linear regression analysis between the dependent LogSEC variable and the explanatory variables NL, CD, T and pH constructed with the validation sample subset (n = 15) showed constancy of variance according to the Modified Levene test conducted with $\alpha = 0.05$ ($^\star t = -0.045 < t_{critical} = 2.298$). The normality of the residuals was confirmed by a joint examination of the normal probability plot and the frequency distribution histogram of the residual classes compared to the normal predicted curve (Figure 2).

Since the variables were collected in a temporal sequence corresponding to the seasonal continuity of the flood and dry periods, it was necessary to analyze the correlation among close errors (residuals). The graphic analysis of the total errors of the regression model, disposed sequentially in time, demonstrated that there was no temporal correlation (Figure 3).
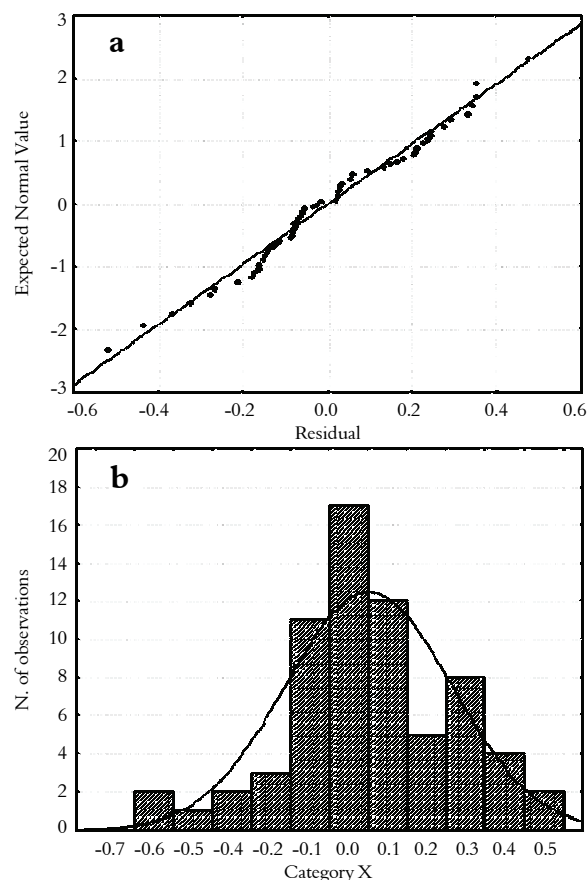

Figure 2. Residual normality: a) normal probability plot, and b) histogram of frequency distribution.
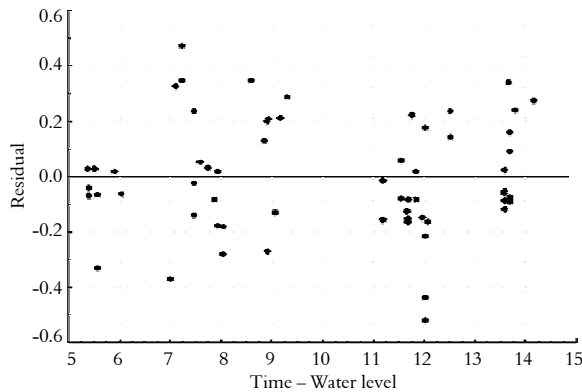
Figure 3. Residual sequential distribution in relation to seasonal variations and the water level.

The presence of values that diverge greatly from the average (outliers) was identified by Standardized Residual and Leverage tests. This analysis put in evidence three outliers in Y and 3 in X, respectively (Figure 4). Using test DFFITS $((DFFITS)_i < 1$ ou $2\sqrt{p/n} = 0.563)$ and the statics of Cook's Distance ($D_i < 10\%$ of the value from $F_{(5; 58)} = 1.696$), we verified that the predicted outlier values, did not significantly influence the estimated regression line.
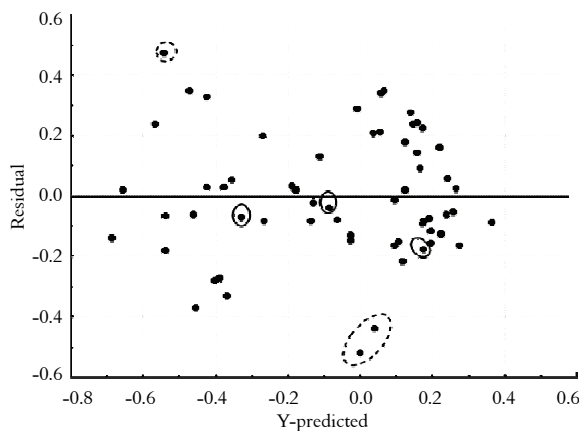
Figure 4. Relation between errors and predicted Y values. Outliers in Y (traced circles) in X (continuous circles).

During the model validation, the β_0, β_1 parameters and its respective standard deviations $s\{\beta\}$ between the regression models originated from training and validation data were compared. In this analysis, the validity of the model generated is a result of the similarity between the parameters. According to the limits reached by β_0 and β for each variable in the training data and validation, there was little difference between β_{NL} and β_0; β_{CD} was reversed but very similar and only β_{TF} and β_{pH} showed large variations (Table 5). However, the large overlap between the coefficients β_0 and $\beta_{1,2,3,4}$ of

training and validation models by the assigned standard deviations validates the model for this analysis.

Table 5. Parameters of regression model for the training and validation data.

Statistics	Training	Validation
b_0	4.360	3.981
$s\{b_0\}$	1.025	2.755
b_1	0.041	0.060
$s\{b_1\}$	0.015	0.021
b_2	-0.085	-0.113
$s\{b_2\}$	0.026	0.036
b_3	0.003	-0.000
$s\{b_3\}$	0.001	0.004
b_4	-0.398	-0.196
$s\{b_4\}$	0.135	0.449
SQRes*	2.537	0.401
QMRes**	0.043	0.040
R^2	0.66	0.77

*SQRes: Sum of square errors. **QMRes: Mean square errors.

As a complement to this validation process, a linear regression analysis was done between Y observed (Y) and Y predicted ($\hat{Y}$), aiming to examine if the regression line passes through the origin and if the correspondence between Y and $\hat{Y}$ is equivalent to one unit (100% of agreement among training and validation data). The t hypothesis test demonstrated that the regression passes through the origin (t★ = 1.676 < $t_{critical}$ = 2.160), but that there was no fitting between the models (t★ = 6.758 > $t_{critical}$ = 2.160, Figure 5).
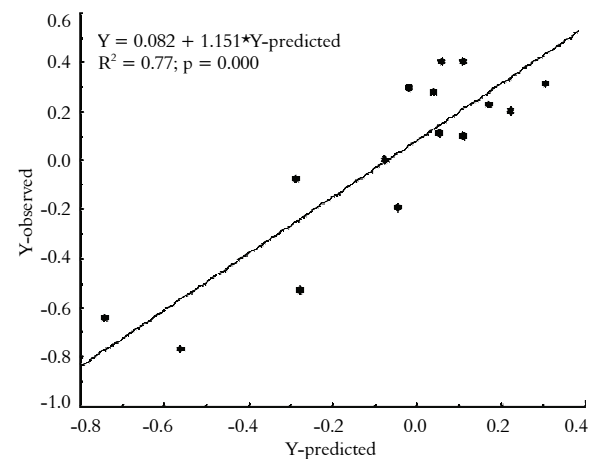
Figure 5. Regression between observed Y and predicted Y, calculated from the parameters of the model considered.

The non-fitting of the model chosen with validation data can be related to the sampling bias due to the small number of validation samples. This causes biased results because it does not totally represent the environmental phenomenon analyzed. Nevertheless, the determination coefficient ($R^2 = 0.77$) and the p-value (p < 0.000) of regression analysis and the similarity of β demonstrate a significant fitting among the data analyzed of the model chosen.

The selection of the model and appropriate variables, as well as how the variables enter the model, are complex tasks that must be performed by taking care to maximize the inter-relationships between variables (RAFTERY; DEAN, 2006). In this study, the results show the potential of the procedures used in developing the model for calculating water transparency in terms of physical and chemical variables selected for the lake ecosystems analyzed.

We suggest considering the temporal continuity during the collection of sampling data, including the transition periods between flood and drought characteristic of the Amazon hydrologic system. We also recommend the inclusion of other variables for the analysis of water transparency variation, searching for a relationship among parameters of the analyzed model with climatic, physiographic and topographic characteristics of the drainage basins.

Conclusion

The best set of variables that explain the water transparency variation in the lakes of systems Solimões and Jarauá at the RDSM were positively water level and conductivity; that is, the higher value of these variables increased the visual disappearance depth of Secchi disk, and negatively the bottom temperature and the pH, the higher bottom temperature and pH, lower water transparency. The chosen model showed a good fit with the selected variables.

Acknowledgements

References

ALCÂNTARA, E. H.; STECH, J. L.; NOVO, E. M. L. M.; SHIMABUKURO, Y. E.; BARBOSA, C. C. F. Turbidity in Amazon floodplain assessed through a spatial regression model applied to fraction images derived from MODIS/Terra. **Transactions and Geoscience and Remote Sensing**, v. 46, n. 10, p. 2895-2905, 2008.

ALCÂNTARA, E. H.; NOVO, E.; STECH, J.; LORENZZETTI, J.; BARBOSA, C.; ASSIREU, A.; SOUZA, A. A contribution to understanding the turbidity behaviour in an Amazon floodplain. **Hydrology and Earth System Sciences**, v. 14, n. 2, p. 351-364, 2010.

AYRES, J. M. **As matas de floodplain do Mamirauá: médio Rio Solimões**. Brasilia: Sociedade Civil Mamirauá, 1995.

ALMEIDA, F. F.; MELO, S. Considerações limnológicas sobre um lago da planície de inundação Amazônica (lago Catalão – Estado do Amazonas, Brasil). **Acta Scientiarum. Biological Sciences**, v. 31, n. 4, p. 387-395, 2009.

BOURGOIN, L. M.; BONNET, M.-P.; MARTINEZ, J.-M.; KOSUTH, P.; COCHONNEAU, G.; MOREIRA-TURCQ, P.; GUYOT, J.-L.; VAUCHEL, P.; FILIZOLA, N.; SEYLER, P. Temporal dynamics of water and sediment exchanges between the Curuaí floodplain and the Amazon river, Brazil. **Journal of Hydrology**, v. 335, n. 1-2, p. 140-156, 2007.

BUKATA, R. P.; JEROME, J. H.; KONDRATYEV, K. Y.; POZDNYAKOV, D. V. **Optical properties and remote sensing of inland and coastal waters**. Florida: CRC Press LCL, 1995.

CARVALHO, F. T.; VELINI, E. D.; CAVENAGHI, A. L.; NEGRISOLI, E.; CASTRO, R. M. Influência da turbidez da água do rio Tietê na ocorrência de plantas aquáticas. **Planta Daninha**, v. 23, n. 2, p. 359-362, 2005.

CHAVES, R.; CAMARGO, M.; QUEIROZ, H.; HERCOS, A. Ritmo de atividade diária de *Osteoglossum bicirrhosum* (Peixes: Osteoglossiformes) em quatro lagos da Reserva de Desenvolvimento Sustentável Mamirauá, AM. **Uakari**, v. 1, n. 1, p. 71-78, 2005.

COSTA, M. Uso de imagens de radas para o cálculo da produção primária de plantas aquáticas nas várzeas da Amazônia. **Acta Amazônica**, v. 35, n. 2, p. 145-154, 2005.

CRISTOFOR, S.; VADINEANU, A.; IGNAT, G.; CIUBUC, C. Factors affecting light penetration in shallow lakes. **Hydrobiologia**, v. 275/276, n. 1, p. 493-498, 1994.

FORSBERG, B. R.; DEVOL, A. H.; RICHEY, J. E.; MARTINELLI, L. A.; SANTOS, H. Factors controlling nutrient concentrations in Amazon floodplain lakes. **Limnology and Oceanography**, v. 33, n. 1, p. 41-56, 1988.

FREIRE, R. H. F.; CALIJURI, M. C.; SANTAELLA, S. T. Longitudinal patterns and variations in water quality in a reservoir in the semiarid region of NE Brazil: responses to hydrological and climatic changes. **Acta Limnologica Brasiliensia**, v. 21, n. 2, p. 251-262, 2009.

HENDERSON, P. A.; ROBERTSON, B. A. On structural complexity and fish diversity in an Amazonian floodplain. In: PADOCH, C.; AYRES, J. M.; PINEDO-VASQUEZ, M.; HENDERSON, P. A. (Ed.). **Varzea: diversity, development, and conservation of Amazonia's withewater floodplain**. New York: The New York Botanical Garden Press, 1999. p. 45-58.

IRION, G.; JUNK, W. J.; MELLO, J. A. S. N. The large central Amazonian river floodplains near Manaus: geological, climatological, hydrological and geomorphological aspects. In: JUNK, W. (Ed.). **The central Amazon floodplain**. Nova York: Springer, 1997. p. 23-46.

JUNK, W. J. **The central Amazon floodplain**: ecology of a pulsating system. Berlin: Ecological Studies, 1997.

KUTNER, M. H.; NACHTSHEIM, C. J.; NETER, J.; LI, W. **Applied linear statistical models**. Chicago: Mc Graw-Hill/Irwin, 2004.

MELO, S.; SOUZA, K. F. Flutuação anual e interanual da riqueza de espécies de desmídeas (Chlorophyta – Conjugatophyceae) em um lago de inundação amazônico de águas pretas (Lago Cutiuaú, Estado do Amazonas, Brasil). **Acta Scientiarum. Biological Sciences**, v. 31, n. 3, p. 235-243, 2009.

MERTES, L. A. K.; DUNNE, T.; MARTINELLI, L. A. Channel-floodplain geomorphology along the Solimões-Amazon river, Brazil. **Geological Society of America Bulletin**, v. 108, n. 9, p. 1089-1107, 1996.

QUEIROZ, H. L. A pesca, as pescarias e os pescadores de Mamirauá. In: QUEIROZ, H. L.; CRAMPTON, W. G. R. (Ed.). **Estratégias para o manejo de recursos pesqueiros em Mamirauá**. Brasília: Sociedade Civil Mamirauá, 1999. p. 37-71.

QUEIROZ, H. L. A reserva de desenvolvimento sustentável Mamirauá. **Estudos Avançados**, v. 19, n. 54, p. 183-203, 2005.

QUEIROZ, H. L. Classification of water bodies based on biotic and abiotic parameters at the várzeas of Mamirauá reserve, Central Amazon. **Uakari**, v. 3, n. 2, p. 19-34, 2007.

RADKE, R. J.; GAUPISCH, A. Effects of phytoplankton-induced turbidity on predation success of piscivorous Eurasian perch (*Perca fluviatilis*): possible implications for fish community structure in lakes. **Naturwissenschaften**, v. 92, n. 2, p. 91-94, 2005.

RAFTERY, A. E.; DEAN, N. Variable Selection for Model-Based Clustering. **Journal of the American Statistical Association**, v. 101, n. 473, p. 168-178, 2006.

RUSSEL, M. A.; WALLING, D. E.; HADGKINSON, R. A. Suspended sediment sources in two small lowland agricultural cachtments in the UK. **Journal of Hydrology**, v. 252, n. 1, p. 1-24, 2001.

SCHMIDT, G. W. Primary production of phytoplankton in the three types of Amazonian waters. II. The limnology of a tropical flood-plain lake in Central Amazonia (lago Castanho). **Amazoniana**, v. 4, n. 2, p. 139-203, 1973.

SIEGEL, S. **Estatística não-paramétrica** (para ciências do comportamento). São Paulo: McGraw-Hill, 1975.

SIOLI, H. The Amazon and its main affluents: hydrology, morphology of the rivers courses, and river types. In: JUNK, W. (Ed.). **The Amazon**: limnology and landscape ecology of a mighty tropical river and its basin. The Netherlands: Junk Publishers, 1984. p. 127-165.

WETZEL, R. G. **Limnology**: lake and river ecosystems. London: Academic Press, 2001.