

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA**

DISSERTAÇÃO DE MESTRADO

**REVISÃO DE MODELOS PROBABILÍSTICOS DE DISTRIBUIÇÃO: UMA
APLICAÇÃO PARA PEIXES MIGRADORES**

Autor

José Ricardo de Souza Barradas

Orientador

Dr. Nelson Ferreira Fontoura

Porto Alegre – RS – Brasil

Março de 2012

SUMÁRIO

AGRADECIMENTOS	3
RESUMO	4
ABSTRACT	6
1. INTRODUÇÃO	8
1.1. Caracterização geral da problemática dos peixes migradores.....	8
1.2. Modelos probabilísticos para distribuição de organismos	9
2. MATERIAL E MÉTODOS	15
2.1. Descrição da Área de Estudo.....	15
2.2. Programa de amostragem.....	16
2.2.1. Matriz amostral	16
3. RESULTADOS	20
4. DISCUSSÃO	45
5. REFERÊNCIAS BIBLIOGRÁFICAS	52

AGRADECIMENTOS

Aos meus pais José Ricardo Pereira Barradas e Margarete Dorneles de Souza Barradas;

Às minhas irmãs Cecília, Sofia, Ana Lucília e Maria Felícia;

Ao meu orientador Nelson Ferreira Fontoura;

Aos meus amigos Bianca, Joana, Katiucia, Thiago Borges e Vinícius;

Aos meus amigos e colegas de laboratório Gianfranco, Lúcia, Thais e Thiago Silveira;

À Pontifícia Universidade Católica do Rio Grande do Sul;

Ao Programa de Pós-Graduação em Zoologia;

E a todos que, de alguma forma, participaram e ajudaram na execução deste trabalho.

RESUMO

Tendo em vista as limitações dos modelos LOGIT, representado pela equação $P = e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)} \cdot (1 + e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)})^{-1}$, onde: P é a probabilidade de ocorrência da espécie (0-1); x_1 , x_2 e x_i são os descritores ambientais de ocorrência; b_0 , b_1 , b_2 e b_i são coeficientes de calibração do modelo, e LOGIT_m, representado pela equação $P = e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))} \cdot (1 + e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))})^{-1}$, onde: P é a probabilidade de ocorrência da espécie (0-1); b_1 e b_2 são coeficientes relativos à altitude e área de bacia, respectivamente; PMF é o ponto de mudança de fase de cada parâmetro, já descritos em literatura, para predição de distribuição probabilística de espécies em situações-limite devido a mecanismos compensatórios consequentes do desenho desses modelos, o objetivo deste trabalho foi propor um novo modelo estatístico para distribuição de espécies de peixes migradores e compará-lo com os modelos citados acima. Foi utilizada a base de dados disponível derivada do projeto PEIXES MIGRADORES E POTENCIAL HIDRELÉTRICO: GESTÃO INTEGRADA DA BACIA URUGUAI (RS/SC), sendo esta composta por 167 pontos distribuídos por todo o território brasileiro da bacia hidrográfica. O modelo proposto neste estudo (Logistic Product – LP) é respresentado pela seguinte equação: $P = (1 - b_0) + b_0 \cdot (1 + e^{(TAXA_{Altitude} \cdot (Altitude - PMF_{Altitude}))})^{-1} \cdot (1 + e^{(-TAXA_{Área de Bacia} \cdot (Área de Bacia - PMF_{Área de Bacia}))})^{-1}$, onde: P é a probabilidade de ocorrência da espécie (0-1); b_0 é uma fração da probabilidade de ocorrência não explicada por qualquer dos descritores; $TAXA$ é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente; PMF é o ponto de mudança de fase de cada parâmetro. Os modelos foram ajustados para quatro espécies migradoras de grande porte da bacia hidrográfica do rio Uruguai: *Leporinus obtusidens* (piava), *Prochilodus lineatus* (grumatã), *Pseudoplatystoma corruscans* (pintado) e *Salminus brasiliensis* (dourado). Os modelos foram analisados conforme (1) suas aderências entre a ocorrência prevista e estimada; (2) formatos dos campos dimensionais da variância residual em função das variáveis altitude e área de bacia, (3) pelos formatos do campos dimensionais de probabilidade de ocorrência em função das variáveis altitude e área de bacia, e (4) através dos critérios estatísticos Akaike Information Criterion (AIC) e Dimensional Stability Index (DSI), sendo o último desenvolvido neste estudo e representado pela seguinte equação: $D.S.I. = (C_v P_1 \cdot C_v P_2 \cdot C_v P_3 \cdot \dots \cdot C_v P_i)^{1/i}$, onde: $C_v P_1$, $C_v P_2$, $C_v P_3$, $C_v P_i$ é o coeficiente de variação de cada parâmetro. Os percentuais de aderência obtidos foram geralmente próximos a 80%, e, de forma geral, o modelo LP apresentou índices de aderência entre 1% e 4% mais baixos que os demais modelos. Os $PMF_{Altitude}$ obtidos com o modelo LP foram mais altos para todas as espécies, sendo de 672 m para *S. brasiliensis*, 516 m para *P. lineatus*, 651 m para *L. obtusidens* e 509 m para *P. corruscans*. Os $PMF_{Área de Bacia}$ foram maiores para o modelo LOGIT_m, exceto para *P. lineatus*, onde os modelos LP e LOGIT_m apresentaram o mesmo valor (101 km²) e para *P. corruscans*, onde o modelo LP apresentou valor de 1623 km² e o modelo LOGIT_m apresentou valor de 709 km². Para os modelos LOGIT e LOGIT_m os campos de probabilidade de ocorrência das espécies apresentou o mesmo comportamento geral, na forma de um plano oblíquo com seção transversal sigmoide. Entretanto, os campos formados pelo somatório dos quadrados dos resíduos (SQR) obtidos com o modelo LOGIT_m para todas as espécies apresentaram comportamento de “calha” de menores valores de variância residual, indicando igual qualidade estatística ara um amplo conjunto de combinações de PMF de altitude e área de bacia, o que resulta em falta de estabilidade dos parâmetros de ajuste. O modelo LP apresentou sempre um ponto de menor valor de SQR em todos os ajustes, mostrando melhor estabilidade da resposta final, porém, o valor mínimo obtido foi sempre um pouco mais elevado que nos demais modelos. Os campos de

probabilidade de ocorrência mostraram que o modelo LP apresenta interação entre as variáveis ambientais mais próxima da realidade biológica que os modelos LOGIT e LOGIT_m, não sendo verificados mecanismos compensatórios como nos modelos LOGIT e LOGIT_m. O modelo LP apresentou os maiores valores de AIC para todas as espécies, sendo 34,3 para *S. brasiliensis*, 35,0 para *P. lineatus*, 35,2 para *L. obtusidens* e 33,7 para *P. corruscans*. Neste critério o modelo apresentou-se menos adequado que os demais para descrever a distribuição das espécies. Com o método DSI, entretanto, o modelo LP obteve os menores valores, exceto para *S. brasiliensis*, onde alcançou resultados ligeiramente mais alto que o modelo LOGIT_m. No presente estudo evidenciamos a similaridade entre os modelos LOGIT e LOGIT_m, pois, com exceção de *L. obtusidens*, as diferenças entre eles foram extremamente pequenas, não sendo representativas para estudos nesta escala. O modelo LP não apresentou mecanismos compensatórios como os modelos LOGIT e LOGIT_m, portanto, embora tenha obtido resultados piores no critério AIC, as consequentes melhorias quanto a estabilidade estatística, formato do campo dimensional e melhor capacidade preditiva em situações-limite justificam sua utilização.

ABSTRACT

Review of probabilistic distribution models: an application to migratory fish

Given the limitations of the logit model, represented by the equation $P = e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)} \cdot (1 + e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)})^{-1}$, where: P is the probability of occurrence of the species (0-1), x_1 , x_2 and x_i are the environmental descriptors, b_0 , b_1 , b_2 and b_i are coefficients of the model calibration, and LOGIT_m, represented by the equation $P = e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))} \cdot (1 + e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))})^{-1}$, where: P is the probability of occurrence of the species (0-1), b_1 and b_2 are coefficients related to altitude and basin area, respectively, and PMF is the point of changing phase of each parameter, as described in the literature for prediction of probability distribution of species in extreme situations due to compensatory mechanisms resulting from the design of these models, the goal of this study was to propose a new statistical model for the distribution of species of migratory fish and compare it with the models mentioned above. It was used the available database derived from the project PEIXES MIGRADORES E POTENCIAL HIDRELÉTRICO: GESTÃO INTEGRADA DA BACIA URUGUAI (RS/SC), which is composed of 167 points distributed throughout the Brazilian territory of the watershed. The model proposed in this study (Logistic Product - LP) is represented by the following equation: $P = (1 - b_0) + b_0 \cdot (1 + e^{(TAXA_{Altitude} \cdot (Altitude - PMF_{Altitude}))})^{-1} \cdot (1 + e^{(-TAXA_{Área de Bacia} \cdot (Área de Bacia - PMF_{Área de Bacia}))})^{-1}$, where: P is the probability of occurrence of the species (0-1), b_0 is a fraction of the likelihood of unexplained by any of the descriptors; $TAXA$ is a parameter indicative of the relative speed of change of absent/presence and PMF is the point of changing phase of each parameter. The models were adjusted for four species of migratory large Uruguay River basin: *Leporinus obtusidens* (piava), *Prochilodus lineatus* (grumatã) *Pseudoplatystoma corruscans* (pintado) and *Salminus brasiliensis* (dourado). The models were analyzed according to (1) adherence between their expected occurrence and estimated, (2) formats dimensional fields of the residual variance as a function of the variables altitude and basin area, (3) the formats dimensional fields of probability of occurrence in function of the variables altitude and basin area, and (4) through statistical criteria of Akaike Information Criterion (AIC) and Dimensional Stability Index (DSI), the latter being developed in this study and represented by the following equation: $D.S.I. = (C_v P_1 \cdot C_v P_2 \cdot C_v P_3 \cdot \dots \cdot C_v P_i)^{1/i}$, where: $C_v P_1$, $C_v P_2$, $C_v P_3$, $C_v P_i$ is the coefficient of variation for each parameter. The percentage of adherence obtained were generally around 80%, and, in general, the LP model showed adherence rates between 1% and 4% lower than the other models. The $PMF_{Altitude}$ obtained with the LP model were higher for all species, being 672 m for *S. brasiliensis*, 516 m for *P. lineatus*, 651 m for *L. obtusidens* and 509 m for *P. corruscans*. The $PMF_{Basin Area}$ were higher for the model LOGIT_m, except for *P. lineatus*, where the models LOGIT_m and LP had the same value (101 km²) and *P. corruscans*, where the LP model showed a value of 1623 km² and LOGIT_m model showed a value of 709 km². The fields of probability of occurrence for the LOGIT and LOGIT_m models for all species had the same general behavior in the form of an oblique cross-section sigmoid. However, the fields formed by the residual sum of squares (RSS) obtained with the model LOGIT_m for all species behaved like a "trough" of lower values of variance, indicating the same statistical quality ara a wide range of combinations of $PMF_{altitude}$ and $PMF_{Basin Area}$, which results in lack of stability of the fitting parameters. The LP model presented only one point of lower value of RSS in all settings, showing better stability of the final answer, however, the minimum value obtained was always a little higher than in other models. The fields of probability of occurrence showed that the LP model shows the interaction between environmental

variables closer to the biological reality that LOGIT and LOGIT_m, not being checked the compensatory mechanisms presented in the models LOGIT and LOGIT_m. The LP model presented the highest values of AIC for all species, and 34.3 for *S. brasiliensis*, 35.0 for *P. lineatus*, 35.2 for *L. obtusidens* and 33.7 for *P. corruscans*. The model presented in this criterion is less suitable than others to describe the distribution of species. With the DSI method, however, the LP model obtained the lowest values, except for *S. brasiliensis*, which achieved results slightly higher than the model LOGIT_m. In the present study we noted the similarity between models LOGIT and LOGIT_m because, except for *L. obtusidens*, the differences between them were extremely small, not representative for the studies on this scale. The LP model showed no compensatory mechanisms such as the LOGIT and LOGIT_m, therefore, although the worst results obtained in the AIC criterion, the resulting improvements in the statistical stability, shape of dimensional field and better predictive ability in extreme situations justify its use.

1. INTRODUÇÃO

1.1. Caracterização geral da problemática dos peixes migradores

Peixes migradores realizam deslocamentos populacionais periódicos ao longo do ciclo de vida, sendo esse comportamento fundamental para manutenção de populações viáveis (GODINHO & KYNARD, 2009). A migração pode se estender por muitos quilômetros – a área de bacia necessária para a sobrevivência de muitas espécies migradoras pode chegar a 40.000km² – mostrando que o metabolismo desses peixes está possivelmente ligado a necessidade biológica de nadar, sendo este um dos processos que estimula os órgãos sexuais e o amadurecimento de gametas (GODOY, 1987; CAPELETI & PETRERE JR, 2006). Em ambientes temperados, muitas espécies compartilham padrões gerais de comportamento. A desova ocorre nas partes mais altas dos rios e a corrente carrega ovos e larvas para locais de alimentação, onde estes se desenvolverão até se unirem ao estoque principal (PITCHER & HART, 1982).

As hidrelétricas têm interferido de forma marcante nos processos migratórios de peixes (HALL, 1972). Para minimizar tal impacto, a Lei nº. 2.250 tornava obrigatória a construção de escadas em barramentos já em 1927 (Centrais Elétricas do Sul do Brasil S/A 1985). Embora muitas escadas de peixes tenham sido construídas, grande parte dos barramentos é desprovida de estações de piscicultura ou caminhos para peixes, e aquelas que os tem, na maioria das vezes são ineficientes, resultando na depleção das populações de montante (BERTOLETTI, 1987; AGOSTINHO *et al.*, 2007). Dada a relevância do problema, em agosto de 2007 o assunto foi debatido pela comunidade científica durante International Symposium of Fish Passage in South America.

Grande parte das hidrelétricas brasileiras localiza-se em rios de planalto ou planície, onde os rios tem grande vazão mas pequena declividade. Neste contexto, os barramentos se constituem em obras de grande envergadura e que afetam diretamente aspectos da biologia dos peixes não só pela interrupção, alteração e fragmentação do hábitat original, mas perda do padrão de fluxo sazonal causado pela menor variação de vazão a jusante da barragem (LOWE-MCCONNELL, 1999; KELMA *et al.*, 2002; ANDRADE & BRAGA, 2005).

Os peixes de água doce desenvolveram mecanismos adaptativos para a sobrevivência em ambientes onde as condições bióticas e abióticas variam amplamente segundo um ciclo sazonal, sendo este ciclo necessário para a maturação gonadal e obtenção de alimento nas diferentes fases do ciclo de vida (VAZZOLER,

1996). Neste sentido, a regularização do fluxo de água através de barramentos mostre-se prejudicial para o desenvolvimento de muitas espécies.

1.2. Modelos probabilísticos para distribuição de organismos

Segundo Waddington (1979), a matemática tem grande utilidade para atenuar dificuldades no entendimento de sistemas biológicos mais complexos, ou mesmo para destacar algumas complexidades não visíveis na natureza. Fenômenos biológicos podem apresentar aspectos que somente são identificados ou previstos por meio de modelos matemáticos (KATO & BELLINI, 2009).

O rápido desenvolvimento de Sistemas de Informação Geográfica (GIS) e a crescente disponibilidade de modelos digitais de terreno têm aumentado a importância dos modelos probabilísticos de distribuição de espécies (GUISAN & ZIMMERMANN, 2000; BRAUNISCH *et al.*, 2008). Estes têm se tornando um componente fundamental na pesquisa científica, sendo utilizados para testar hipóteses biogeográficas e ecológicas, principalmente na investigação das relações espécie-ambiente (AUSTIN, 2002; GUISAN & THUILLER, 2005; AUSTIN, 2007; FRANKLIN, 2011), provendo não só um melhor entendimento dessa relação, mas também uma predição da distribuição das espécies, mesmo em locais não amostrados (ALVES & FONTOURA, 2009; MOORE *et al.*, 2009).

De forma geral, a modelagem probabilística tem ganhado valor como ferramenta para planejamento e conservação ambiental (GUISAN & THUILLER, 2005; PEARSON, 2007), sendo utilizada como ferramenta para prever os impactos das mudanças ambientais na distribuição dos organismos (LISCHKE *et al.*, 1998; CLARK *et al.*, 2001), para testar hipóteses biogeográficas (MOURELL & EZCURRA, 1996), para planejamento ambiental (MOILANEN *et al.*, 2008), ou mesmo para identificar prioridades de conservação (MARGULES & AUSTIN, 1994; PEARCE & FERRIER, 2000a; ELITH *et al.*, 2011). Franklin (2011) considera os modelos de distribuição de espécie como uma das mais importantes formas de modelagem, porque é capaz de extrapolar distribuição das espécies no espaço e no tempo (FRANKLIN, 2010).

Para ser utilizado, o modelo deve permitir a estimativa precisa da probabilidade de ocorrência de uma determinada espécie, considerando fatores como a natureza exata da relação entre a distribuição e cada variável ambiental com capacidade preditiva (PEARCE & FERRIER, 2000b).

O modelo LOGIT (equação 1, abaixo) foi inicialmente empregado por Joseph Berkson para a resolução de problemas biológicos, e tem sido empregado sistematicamente desde o fim da década de 1950 (CRAMER, 2003).

$$P = e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)} \cdot (1 + e^{(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i)})^{-1} \quad (1)$$

Onde:

P é a probabilidade de ocorrência da espécie (0-1);

x_1, x_2 e x_i são os descritores (parâmetros) ambientais de ocorrência;

b_0, b_1, b_2 e b_i são coeficientes de calibração do modelo.

Modelos de regressão logística tornam-se adequados para este tipo de modelagem, pois permitem uma predição probabilística da ocorrência dos organismos e trabalham com a variável dependente como uma variável categórica de dois estados, como por exemplo, presente/ausente (PORTER *et al.*, 2000; BEWICK *et al.*, 2005; DTREG, 2008). É esse o ponto que distingue um modelo de regressão logística de um modelo de regressão linear: a interpretação do resultado da regressão logística é binária ou dicotômica, mesmo embora o resultado numérico seja contínuo entre zero e um. Essa diferença entre regressão logística e linear se reflete tanto na escolha do modelo quanto na hipótese a ser testada. Uma vez que esta diferença é contabilizada, os métodos utilizados na análise por regressão logística seguem os mesmos princípios gerais utilizados na regressão linear (HOSMER & LEMESHOW, 2000). Diferentemente da maioria dos modelos multivariados, a regressão logística não necessita que as variáveis tenham distribuição normal e sejam relacionadas linearmente (PORTER *et al.*, 2000).

Porter e colaboradores (2000) empregaram o modelo LOGIT para prever a ocorrência de peixes na Columbia Britânica utilizando um conjunto de 14 descritores ambientais. Diversos fatores ecológicos podem potencialmente influenciar a distribuição de peixes, tal como temperatura, pH, alcalinidade, dureza, condutividade elétrica, turbidez, oxigênio dissolvido, velocidade de corrente, etc. Segundo Espínola *et al.* (2009), quatro variáveis ambientais foram necessárias para prever a ocorrência de espécies invasoras em represas na bacia do rio Paraná: temperatura, transparência da água, profundidade do lago e área da represa. Entretanto, a obtenção de dados diretos de qualidade de água representa esforços consideráveis em termos de custo e logística, sendo normalmente indisponíveis para estudos em grandes áreas.

Em larga escala, a utilização de fatores simples derivados de Sistemas de Informação Geográfica (GIS) parecem ser suficientes para o entendimento dos padrões de distribuição de espécies ao longo de uma bacia hidrográfica (MATTHEWS *et al.*, 1992; POUILLY *et al.*, 2006; ALVES & FONTOURA, 2009). Utilizando apenas

altitude e área de bacia como descritores ambientais indiretos, Alves & Fontoura (2009) foram capazes de prever a ocorrência de peixes migratórios na bacia Jacuí com aderências superiores a 80% em relação a ocorrência verificada.

Gonçalves (2010) modificou o modelo LOGIT com vistas a permitir a interpretação direta dos coeficientes de calibração, os quais passam a indicar os valores-limite de cada descritor ambiental em termos de presença ou ausência das espécies (Equação 2). Dois parâmetros foram incorporados: ponto de mudança de fase de altitude e de área de bacia. Esses parâmetros expressam diretamente, e em escala real, o ponto onde as espécies apresentam 50% de probabilidade de ocorrência para cada uma das variáveis independentes quando analisadas isoladamente.

$$P = \frac{e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))}}{1 + e^{(b_1 \cdot (Altitude - PMF_{Altitude}) + b_2 \cdot (Área de Bacia - PMF_{Área de Bacia}))}} \quad (2)$$

Onde:

P é a probabilidade de ocorrência da espécie (0-1);

b_1 e b_2 são coeficientes relativos à altitude e área de bacia, respectivamente, indicando a velocidade relativa de mudança de estado ausente/presente.

PMF é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente.

A utilização de apenas duas variáveis independentes no modelo proposto por Gonçalves (2010) torna problemático o correto ajuste dos parâmetros. Quando testado em uma matriz com todos os valores de altitude e área de bacia potencialmente presentes, o campo da variância residual apresentou-se como um “vale” de menor variância residual (Figura 1). Ou seja, é possível alcançar, com diferentes combinações de pontos de mudança de fase de altitude e de área de bacia, pontos muito similares em termos de menor variância residual, o que problematiza o processo de ajuste. Sendo assim, embora se consiga um bom ajuste estatístico do modelo, o resultado não apresenta estabilidade de convergência em função dos parâmetros-semente inicialmente empregados no processo de ajuste, visto que existem inúmeras soluções de qualidade estatística equivalente, sugerindo a necessidade o desenvolvimento de modelos matemáticos alternativos.

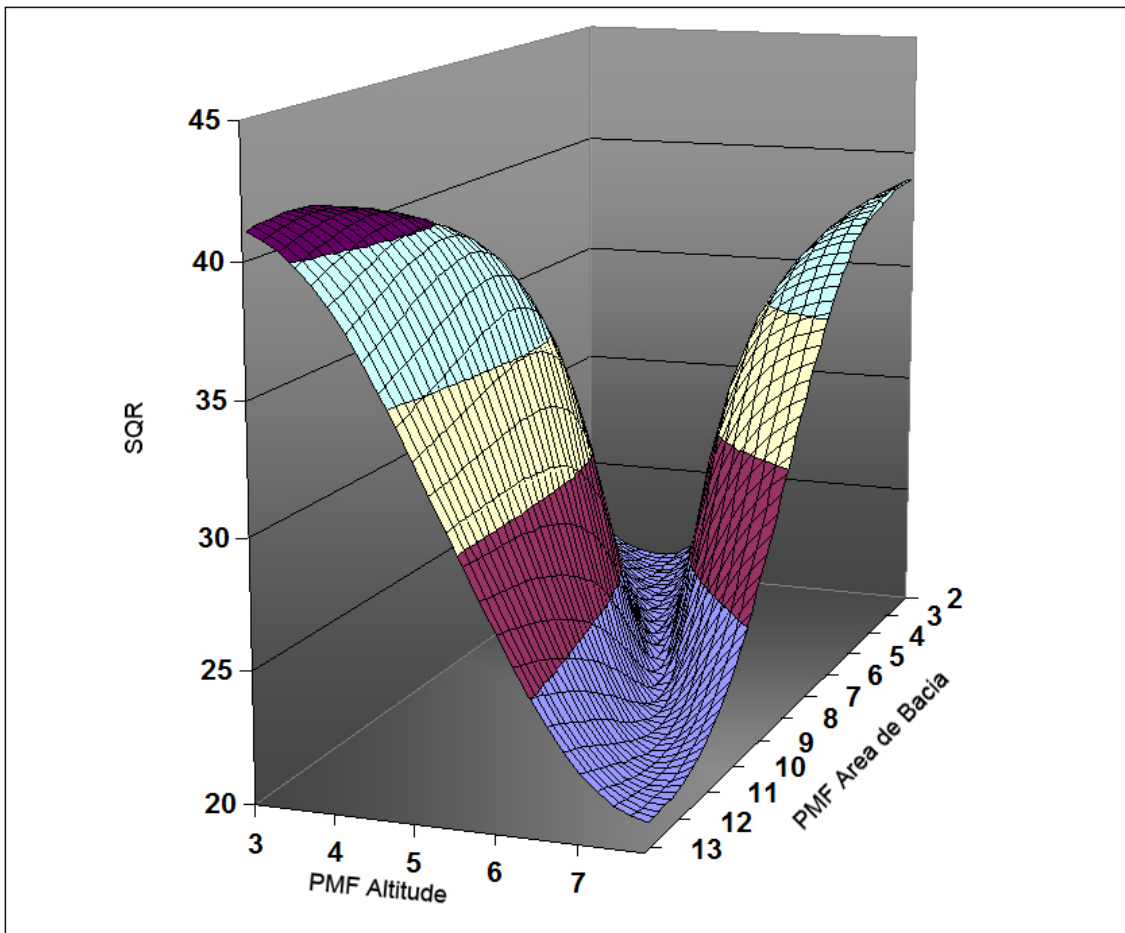


Figura 1: Campo formado pelo somatório do quadrado dos resíduos através do modelo LOGIT_m conforme as combinações de todos os valores possíveis de Pontos de Mudança de Fase de altitude e área de bacia. Dados de *Salminus brasiliensis* para a bacia do rio Uruguai segundo modelo proposto por Gonçalves (2010). SQR = Soma do Quadrado dos Resíduos; PMF = Ponto de Mudança de Fase. As variáveis independentes são representadas em escala logarítmica (ln) para PMF de área de bacia (km²) e altitude (m).

Atualmente existem diversos métodos e softwares dedicados para predição de ocorrência de espécies. Entretanto, muitos destes softwares utilizam apenas dados de presença para estimar a distribuição de organismos, como por exemplo, GARP e Maxent, sendo ideais para modelagens através de dados de museus, coleções, etc, onde o dado de ausência não está disponível (STOCKWELL & PETERS, 1999; ANDERSON *et al.*, 2003; PHILLIPS *et al.*, 2006; ELITH *et al.*, 2011).

Métodos como estes criam pontos de “pseudo-ausência” para ajustar o modelo, conseguindo assim a informação binária de presença e ausência (1-0). Para obter a informação de “pseudo-ausência”, o software seleciona uma parcela dos pontos não-amostrados e os considera como sendo de “ausência”. No caso do software GARP, o dado de ausência é obtido através de testes entre os valores das variáveis

independentes dos pontos não-amostrados contra os valores destas variáveis dos pontos amostrados. A matriz de presença é analisada pelo programa, e este gera um intervalo para cada variável independente com seus valores máximos e mínimos. Os dados das variáveis independentes dos pontos não-amostrados são então comparados com estes intervalos e, caso alguma variável tenha seu valor não contido neste intervalo, o ponto é considerado como de “ausência” (STOCKWELL & PETERS, 1999; PHILLIPS *et al.*, 2006).

Embora seja possível modelar a distribuição de organismos apenas com dados de presença, estes métodos devem ser utilizados como alternativa para casos onde os dados de presença/ausência não estão disponíveis (ELITH *et al.*, 2011). Variações do esforço amostral causam maior interferência sobre modelagens com dados de apenas presença, portanto o resultado dessas modelagens não é somente a distribuição das espécies, e sim uma combinação disto com a distribuição do esforço amostral (PHILLIPS *et al.*, 2009; ELITH *et al.*, 2011).

Métodos de Redes Neurais Artificiais também tem sido aplicados em modelagem ecológica para prever a ocorrência de espécies onde há dados de presença e ausência. Neste caso, cria-se uma rede de várias camadas onde os elementos de processamento da rede em cada uma das camadas (neurônios), são discretos no tempo. A primeira camada (camada de entrada) conecta-se às variáveis ambientais (variáveis de entrada), onde um neurônio representa cada uma das variáveis. A última camada (camada de saída) conecta-se às variáveis de saída, onde o número de neurônios é igual ao número de categorias nas quais os indivíduos estão sendo classificados. Para modelagem de distribuição usa-se dois neurônios na camada de saída, um para presença e um para ausência. Entre as camadas de entrada e saída pode haver uma ou mais camadas (camadas ocultas), onde todos os neurônios de uma camada são ligados com todos os neurônios das camadas adjacentes. O número de camadas ocultas é um importante parâmetro da rede e é escolhido através de múltiplos testes empíricos de várias combinações, sendo escolhida a configuração que melhor represente o problema estudado (KOHAVI, 1995; MASTRORILLO *et al.*, 1997).

Embora venha sendo cada vez mais utilizado, há uma falta de poder explicativo no método de Redes Neurais Artificiais, pois não é possível interpretar, seja de forma teórica ou prática, os pesos das variáveis ambientais (MASTRORILLO *et al.*, 1997).

Um dos problemas relativos ao uso de modelagem matemática para previsão de ocorrência relaciona-se ao balanço entre a complexidade do modelo e o seu poder

de descrição. Ao mesmo tempo que modelos mais complexos podem ser mais precisos, o custo adicional de novos descritores pode ser indesejável.

O desenvolvimento de diversos modelos para predição de organismos tornou necessária a utilização de indicadores estatísticos para selecionar o modelo que melhor se adéque ao problema estudado. Através do uso desses indicadores pode-se identificar, além da adequabilidade do modelo, se a complexidade deste é justificada em termos de ganho de aderência ao problema biológico (MASTRORILLO *et al.*, 1997). A seleção estatística de modelos tem sido amplamente utilizada em problemas de ecologia e evolução, particularmente, a seleção de modelos é um pré-requisito para a maioria dos estudos filogenéticos e de marcação-recaptura (LEBRETON *et al.*, 1992; JOHNSON & OMLAND, 2004).

A tabela 1 apresenta cinco métodos frequentemente utilizados em estudos de ecologia e evolução. O teste de máxima verossimilhança (LRT) compara pares de modelos aninhados, onde um modelo mais simples é derivado de um modelo mais complexo. Quando a verossimilhança do modelo mais complexo é significativamente maior que a do modelo mais simples, este é mais adequado, ou seja, o custo da adição de parâmetros é compensado pela representatividade do modelo. O método de Akaike (Akaike Information Criterion, ou AIC) estima a perda da informação de Kullback-Leibler de cada modelo, de maneira que o melhor modelo é aquele que perde menos informação. Este método ainda possui uma variação, denominada de Small Sample Unbiased AIC (AIC_c), que possui fatores de correção para amostras de pequeno tamanho e que deve ser utilizado quando o número de parâmetros exceda $\sim n/40$ (onde n é o tamanho da amostra) (BURNHAM & ANDERSON, 2004; JOHNSON & OMLAND, 2004). O método de Schwarz (Schwarz Criterion, ou SC, também chamado de Bayesian Information Criterion – BIC) é estruturalmente parecido com o método AIC, porém adiciona um parâmetro relativo ao tamanho da amostra. Consequentemente, SC tende a favorecer modelos mais simples, especialmente em amostras grandes (ZUCCHINI, 2000). O método AIC geralmente é mais utilizado por ser apoiado na teoria da informação de Kullback-Leibler, entretanto, alguns autores preferem o método SC porque este tende a mostrar como melhores os modelos menores e mais simples. Testes como o LRT tem claras limitações, e, entre os testes que consideram adequabilidade e complexidade, pesquisadores tem trocado o LRT por métodos como o AIC e SC (JOHNSON & OMLAND, 2004).

Tabela 1: Critérios estatísticos frequentemente empregados para seleção de modelos matemáticos (adaptado de JOHNSON & OMLAND, 2004). RSS é a soma dos quadrados dos resíduos; n é o tamanho da amostra; p é o número de parâmetros; s é o dado referente ao modelo mais simples, f é o dado referente ao modelo mais complexo; y são os dados.

Método	Equação	Teste
Adjusted R^2	$R^2_{adj} = (1 - RSS/n - p - 1) \cdot (\sum(y_i - \bar{y})^2/n - 1)^{-1}$	Adequabilidade
Likelihood Ratio Test	$LRT = -2 \cdot (-n/2 \cdot \ln \cdot (RSS_s/n)) - (-n/2 \cdot \ln \cdot (RSS_f/n))$	Adequabilidade e complexidade
Akaike Information Criterion	$AIC = -2 \cdot (-n/2 \cdot \ln \cdot (RSS/n)) + 2p$	Adequabilidade e complexidade
Small sample unbiased AIC	$AIC_c = -2 \cdot (-n/2 \cdot \ln \cdot (RSS/n)) + 2p \cdot (n/(n-p-1))$	Adequabilidade e complexidade
Schwarz Criterion	$SC = -2 \cdot (-n/2 \cdot \ln \cdot (RSS/n)) + p \cdot \ln(n)$	Adequabilidade, complexidade e tamanho da amostra

2. MATERIAL E MÉTODOS

2.1. Descrição da Área de Estudo

O Estado do Rio Grande do Sul é drenado por densa malha hidrográfica. Os rios, em sua maioria de planalto com grande quantidade de corredeiras e quedas d'água, encontram-se em um clima úmido, possuindo elevado potencial energético, em grande parte já utilizado (JUSTUS, 1990).

O rio Uruguai é formado da confluência do rio Pelotas com o rio Canoas, apresentando um percurso de 2.200km de extensão, até a sua foz no estuário do rio da Prata (Argentina). A área total drenada pelo rio é de cerca de 385.000km², destes, 45% (ou seja, 174.412km²) estão situados em território brasileiro. Em uma área de 76.209km², apresenta relevo com grande declividade, com cotas variando de cerca de 1.600m a 160m acima do nível do mar. O clima da região hidrográfica do Uruguai é temperado, apresentando uma distribuição intra-anual de chuvas relativamente homogênea, porém com alguma elevação no período de maio a setembro, coincidindo com o inverno. Segundo o Documento Base de Referência do Plano Nacional de Recursos Hídricos, a precipitação média anual é de 1.784mm, com temperatura média

anual variando entre 16 e 20°C, e evapotranspiração média anual de 1.041mm. Caracterizado por ser um dos mais importantes corredores de biodiversidade do Cone Sul, o rio Uruguai apresenta em sua fauna diversas espécies endêmicas ou em vias de extinção (PAIM & ORTIZ, 2006), sendo que em território brasileiro a bacia possui remanescentes de 17% do total de sua cobertura florestal original.

2.2. Programa de amostragem

2.2.1. Matriz amostral

Foi utilizada uma base de dados disponível derivada do projeto PEIXES MIGRADORES E POTENCIAL HIDRELÉTRICO: GESTÃO INTEGRADA DA BACIA URUGUAI (RS/SC). A matriz amostral contém 167 pontos para a bacia do rio Uruguai. A amostragem foi feita através do levantamento de dados do sistema SIBIP/NEODAT III e de saídas a campo. As expedições foram efetuadas aproveitando-se a malha rodoviária e seus pontos de cruzamento com os rios em análise, onde foram efetuadas entrevistas com moradores/pescadores locais (Figura 2).

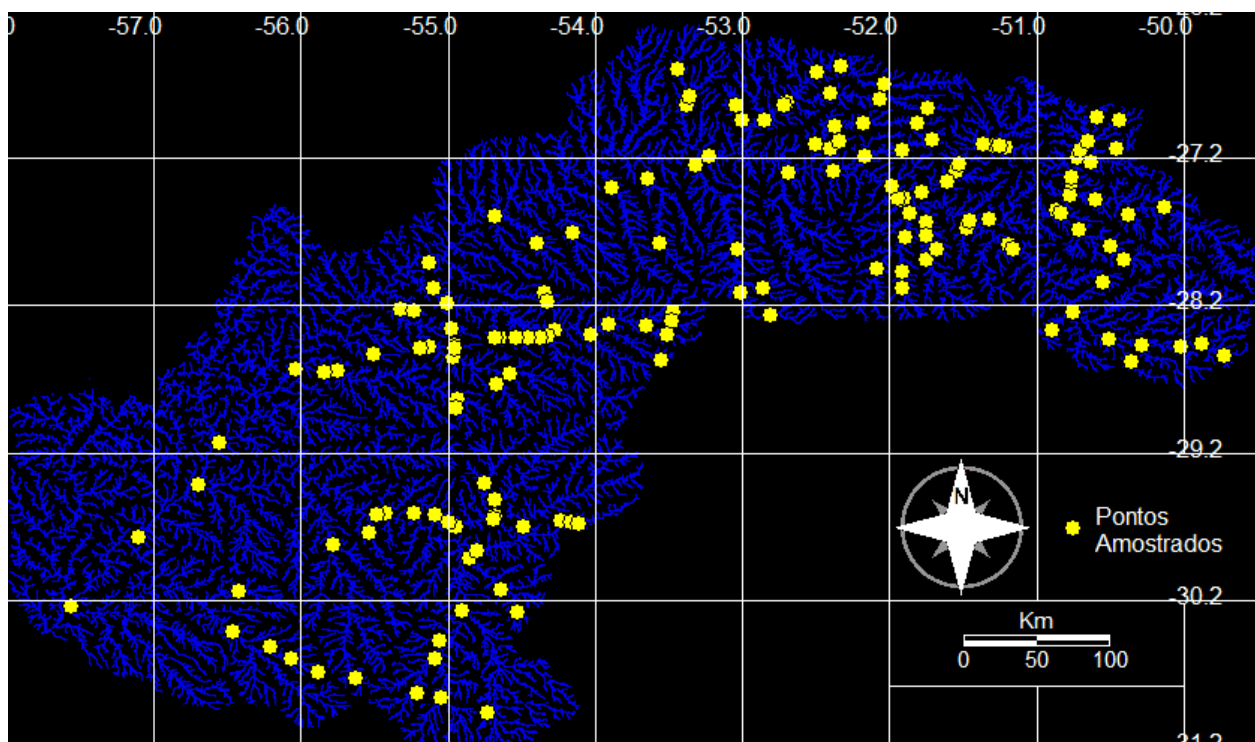


Figura 2. Matriz de 167 pontos amostrados na bacia hidrográfica do rio Uruguai (modificado de GONÇALVES, 2010).

2.2.2 Processamento dos dados

As variáveis mensuradas, assim como a escala de medida, apresentam-se na tabela abaixo:

Tabela 2: Variáveis mensuradas e escalas de medida

<i>Variável</i>	<i>Escala de Medida</i>	<i>Origem do Dado</i>
Presença (por espécie)	Binária: presente (1); ausente (0)	Base de dados disponível
Altitude	Métrica (logaritmo natural da altitude em metros)	Modelo numérico do terreno (Weber <i>et al.</i> 2004)
Área da sub-bacia de montante	Métrica (logaritmo natural da área de bacia em km ²)	Modelo numérico do terreno (Weber <i>et al.</i> 2004; modelado a partir da matriz de altitude utilizando-se as funções <i>Filter Min 3x3</i> , <i>Pit-Removal</i> e <i>Runoff</i>)

Além dos modelos já utilizados por Alves & Fontoura (2009) e Gonçalves (2010) (equações 1 e 2 citadas nas páginas 9 e 11), desenvolvemos e testamos o seguinte modelo estatístico para distribuição de peixes:

Modelo Logistic Product (LP):

$$P = (1 - b_0) + b_0 \cdot (1 + e^{(TAXA \cdot Altitude \cdot (Altitude - PMF_{Altitude}))})^{-1} \cdot (1 + e^{(-TAXA \cdot \text{Área de Bacia} \cdot (\text{Área de Bacia} - PMF_{\text{Área de Bacia}}))})^{-1} \quad (3)$$

Onde:

P é a probabilidade de ocorrência da espécie (0-1);

b_0 é uma fração da probabilidade de ocorrência não explicada por qualquer dos descritores;

$TAXA$ é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente;

PMF é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente.

Os produtos cartográficos utilizados no presente trabalho foram elaborados através do software Idrisi Andes 15.0 (CLARK LABS, 2006), com base em um Modelo Digital de Terreno (MDT, altitude radar em pixel de 92 por 92m) adaptado para o sistema de referência oficial brasileiro (SAD69) (WEBER *et al.*, 2004). As imagens que serviram como base para gerar os mapas deste trabalho foram extraídas do banco de dados do projeto PEIXES MIGRADORES E POTENCIAL HIDRELÉTRICO: GESTÃO INTEGRADA DA BACIA URUGUAI (RS/SC). A hidrografia total amostrada que serviu como máscara para todo o processo de análise foi obtida através de quatro passos básicos: (1) foi realizada a homogeneização da imagem do MDT com a aplicação do *Filter Min 3x3*; (2) estabelecimento de caminhos com elevações monotonicamente decrescentes através da opção *Pit Removal*; (3) aplicação da função *Runoff* para estimativa da área de bacia de montante para cada pixel, corrigindo-se o número de pixels para área real; (4) e por fim aplicação da rotina *Pathway* de caminhos múltiplos a partir de todas as coordenadas dos pontos amostrais, utilizando-se como matriz de custo o inverso da matriz *Runoff*, resultando em uma imagem binarizada (1-0) descritiva do curso natural da hidrografia.

2.2.3. Ajuste e Comparação dos Modelos

Os modelos foram ajustados através de algoritmos dedicados envolvendo seleção assistida de valores-semente e minimização de variância através da rotina Solver do software Microsoft Excel 2007. Os descritores do modelo probabilístico de ocorrência (b_0 , b_1 , b_2 , *PMF*, *TAXA*) foram calculados através da média simples de 100 repetições utilizando 60% das amostras ($n = 100$) selecionadas aleatoriamente, sendo o erro padrão estimado como o desvio padrão dos resultados obtidos após retiradas manual de *outliers*. A significância de cada parâmetro foi estimada através da estatística Wald (TABACHNICK & FIDEL, 1996) e o ponto de corte de presença/ausência das espécies foi de 0,5 (PEARCE & FERRIER, 2000b; ALVES & FONTOURA, 2009).

A análise de adequabilidade dos modelos foi efetuada através da (1) aderência entre a ocorrência prevista e estimada; (2) formato do campo dimensional da variância residual em função das variáveis altitude e área de bacia, (3) pelo formato do campo dimensional de probabilidade de ocorrência em função das variáveis altitude e área de bacia, e (4) através dos seguintes critérios estatísticos:

Akaike Information Criterion (AIC)

O AIC avalia a adequabilidade dos modelos conforme a qualidade estatística do ajuste, mensurada através de qui-quadrado (χ^2), penalizando o incremento do número de parâmetros, ou seja, quanto mais próximo de zero for o índice alcançado, melhor é o modelo. A equação do AIC foi modificada para ser aplicada neste estudo. Hair *et al.* (1995) descrevem o teste AIC da seguinte forma:

$$AIC = \chi^2 + 2 \cdot p \quad (4)$$

Onde:

χ^2 é o valor de qui-quadrado;

p é o número de parâmetros

As razões para a modificação da equação serão tratadas com maiores detalhes no capítulo 4 desta dissertação. O critério proposto é representado por:

$$AIC_m = SQR + 2 \cdot p \quad (5)$$

Onde:

SQR é a soma do quadrado dos resíduos;

p é o número de parâmetros.

Dimensional Stability Index (DSI)

O *Dimensional Stability Index (DSI)*, desenvolvido neste estudo, avalia o modelo penalizando uma maior variabilidade estimada para o conjunto de descritores ajustados. Quanto menor o valor alcançado, melhor é o modelo.

$$D.S.I. = (C_vP_1 \cdot C_vP_2 \cdot C_vP_3 \cdot \dots \cdot C_vP_i)^{1/i} \quad (6)$$

Onde:

$C_vP_1, C_vP_2, C_vP_3, C_vP_i$ é o coeficiente de variação de cada parâmetro.

3. RESULTADOS

Os modelos foram ajustados para quatro espécies de peixes migradores de grande porte presentes na bacia do rio Uruguai: *Leporinus obtusidens* (piava); *Prochilodus lineatus* (grumatã); *Pseudoplatystoma corruscans* (surubim-pintado); *Salminus brasiliensis* (dourado).

Os coeficientes de ajuste para cada modelo estão representados nas tabelas de 3 a 6 para cada uma das espécies em análise. Considerando que os dados de altitude e área de bacia foram logaritimizados (ln) para processamento, os *Pontos de Mudança de Fase* são também apresentados nas tabelas de 3 a 6 convertidos para a escala métrica (e^{PMF}).

As aderências indicam o percentual de acerto do modelo, sendo: *Aderência Total* um indicativo do percentual de acerto geral do modelo, *Aderência Positiva* um indicativo do percentual de acerto do modelo para presença e *Aderência Negativa* um indicativo do percentual de acerto do modelo para ausência. Os percentuais obtidos pelos modelos foram geralmente próximos a 80% (Tabela 7), mostrando que o uso de apenas dois descritores ambientais (altitude e área de bacia) foi capaz de descrever o padrão de distribuição de peixes migradores com relativa precisão. Em geral, o modelo LP apresentou índices de aderência total entre 1% e 4% mais baixos que o modelo LOGIT. Os percentuais de aderência total, aderência positiva e negativa foram também muitos próximos entre os modelos LOGIT e LOGIT_m para as espécies em análise, exceto para *Leporinus obtusidens*, a qual apresentou uma menor aderência negativa para o modelo LOGIT_m (65%), a qual também repercutiu sobre a aderência total.

Os Pontos de Mudança de Fase indicam o ponto médio da curva relativa a cada variável ambiental onde, caso o parâmetro seja analisado isoladamente, a espécie apresenta probabilidade de ocorrência igual a 50%. Os $\text{PMF}_{\text{Altitude}}$ obtidos com o modelo LP foram, para todas as espécies, mais elevados que os obtidos com o modelo LOGIT_m. *Salminus brasiliensis* foi a espécie que obteve os valores mais altos, apresentando $\text{PMF}_{\text{Altitude}}$ igual a 672 metros ($e^{6.5}$) com o modelo LP, quase três vezes mais alto que o obtido com o modelo LOGIT_m (231 metros, $e^{5.3}$). *Pseudoplatystoma corruscans* foi a espécie que apresentou a maior diferença entre os $\text{PMF}_{\text{Altitude}}$, alcançando valores de 68 metros ($e^{4.2}$) com o modelo LOGIT_m e 509 metros ($e^{6.2}$) para o modelo LP, aproximadamente sete vezes maior. *Prochilodus lineatus* apresentou $\text{PMF}_{\text{Altitude}}$ de 151 metros ($e^{5.0}$) e 516 metros ($e^{6.3}$) para os modelos LOGIT_m e LP, respectivamente. *Leporinus obtusidens* obteve valores de 231 metros ($e^{5.4}$) para o modelo LOGIT_m e 651 metros ($e^{6.5}$) para o modelo LP.

Os $PMF_{\text{Área de Bacia}}$ foram maiores para o modelo $LOGIT_m$, com exceção da espécie *Prochilodus lineatus*, onde os modelos $LOGIT_m$ e LP apresentaram o mesmo resultado (101 km^2 , $e^{4,6}$) e *Pseudoplatystoma corruscans*, que foi a espécie que apresentou valores mais elevados para este parâmetro, 709 km^2 ($e^{6,5642}$) para o $LOGIT_m$ e 1623 km^2 ($e^{7,4}$) para o LP, caracterizando a espécie como presente em áreas mais de jusante em relação as demais. O oposto ocorre com *Leporinus obtusidens*, que apresentou os menores valores de $PMF_{\text{Área de Bacia}}$, 177 km^2 ($e^{5,2}$) para o modelo $LOGIT_m$ e 83 km^2 ($e^{4,4}$) para o modelo LP, sendo esta a espécie que alcança os menores segmentos de rios da bacia. *Salminus brasiliensis* teve os $PMF_{\text{Área de Bacia}}$ iguais a 320 km^2 ($e^{5,8}$) com o modelo $LOGIT_m$ e 166 km^2 ($e^{5,1}$) com o modelo LP.

As figuras 3 a 6 apresentam os campos formados pelo somatório do quadrado dos resíduos (letras “A”, “C” e “E”) e de probabilidade de ocorrência (letras “B”, “D” e “F”) para cada espécie conforme os modelos em análise. As imagens são formadas através da utilização de todas as combinações possíveis de altitude e área de bacia presentes na bacia Uruguai, de forma que todos os trechos de rios são potencialmente simulados.

Para os modelos LOGIT e $LOGIT_m$, o campo dimensional de probabilidade de ocorrência apresentou o mesmo comportamento geral, na forma de um plano oblíquo com seção transversal sigmoide, onde, de certa forma, existe um mecanismo compensatório de transferência de efeito através da interação de ambas as variáveis. Identifica-se, entretanto, diferenças de resposta relativas a cada parâmetro ambiental entre os dois modelos, o que resultou no deslocamento do campo sobre os eixos de altitude e área de bacia. Além disso, os modelos apresentaram altas probabilidades de ocorrência em locais com valores de área de bacia extremamente pequenos (Figuras 3 – 6; imagens “B” e “D”), sugerindo inadequação dos modelos em situações-limite, mas que normalmente encontram-se fora do universo real de combinações possíveis.

Do ponto de vista da estabilidade de ajuste através de algoritmos numéricos, é conveniente que o espaço dimensional das variâncias residuais ou do somatório do quadrado dos resíduos (SQR) apresente apenas um ponto de mínimo, com comportamento tridimensional semelhante a uma fossa de menor variância, como na figura 3, imagens A e E, por exemplo.

Entretanto, os campos formados pelo somatório do quadrado dos resíduos (SQR) obtidos no modelo $LOGIT_m$, para todas as espécies, apresentaram sempre um comportamento de “calha”, indicando igual qualidade estatística de ajuste para um amplo conjunto de combinações dos PMF para área de bacia e altitude, o que resulta

em falta de estabilidade dos parâmetros de ajuste em função dos valores-semente inicialmente empregados. Este comportamento torna evidente a deficiência do modelo para solucionar problemas que envolvam apenas duas variáveis independentes.

O modelo LP apresentou somente um ponto de menor valor de SQR em todos os ajustes, indicando estabilidade da resposta final. Porém, o valor mínimo obtido para o SQR foi sempre um pouco mais elevado que nos demais modelos (Figuras 3 – 6, imagens “E”), ou seja, o modelo apresenta estabilidade de resposta mas menor qualidade estatística. Por outro lado, as imagens “F” (Figuras 3 – 6) mostram que a interação entre as variáveis ambientais foi mais próxima da realidade biológica no modelo LP que nos modelos LOGIT e LOGIT_m, na medida em que uma variável isolada, como altitude muito alta ou área de bacia muito baixa, pode zerar a probabilidade de ocorrência de uma espécie independentemente do valor das demais variáveis. Desta forma, não se verificam mecanismos compensatório de interação tal como descrito para os modelos LOGIT e LOGIT_m.

O método utilizado para gerar as imagens do modelo LOGIT foi diferente do utilizado nos demais modelos, pois este não apresenta Pontos de Mudança de Fase (PMF). As variáveis independentes nas imagens “A” (Figuras 3 – 6) são, então, os coeficientes b_1 e b_2 , que são diretamente ligados aos parâmetros ambientais. O modelo LOGIT formou em todos os casos um campo com apenas um valor mínimo de SQR, embora este tenha sido mais evidente no ajuste de *Salminus brasiliensis*.

Considerando a qualidade estatística do ajuste (Tabela 7), o modelo LP apresentou os maiores valores de AIC para todas as espécies, alcançando 34,3 para *Salminus brasiliensis*; 35,0 para *Prochilodus lineatus*; 35,2 para *Leporinus obtusidens* e 33,7 para *Pseudoplatystoma corruscans*. Neste critério o modelo apresentou-se menos adequado que os modelos LOGIT e LOGIT_m para descrever a distribuição das espécies. Entretanto, analisando o resultado obtido com o método DSI, o modelo LP obteve os menores valores, exceto para *S. brasiliensis*, onde alcançou resultado ligeiramente mais alto que o modelo LOGIT_m. Os resultados de variância residual seguem a mesma tendência para todas as espécies, onde o melhor resultado foi obtido pelo modelo LOGIT, seguido pelo modelo LOGIT_m e então pelo modelo LP (Tabela 7).

As figuras 7 a 18 apresentam o padrão de distribuição das espécies analisadas conforme cada modelo para toda a bacia hidrográfica do rio Uruguai. Para fins de comparação, apresenta-se igualmente a diferença de probabilidade estimada entre os modelos LOGIT e LOGIT_m assim como entre o modelo LOGIT e o modelo LP,

indicando os trechos da bacia hidrográfica onde houve divergência nos resultados. O ponto de corte utilizado para definir os estados de ausente/presente foi de 0,5.

Embora o modelo LOGIT tenha apresentado um comportamento diferente do modelo LOGIT_m quanto aos campos dimensionais de SQR's, a tendência observada nas figuras 3 – 6 nos campos dimensionais de probabilidade de ocorrência entre estes dois modelos foi novamente encontrada nos produtos cartográficos gerados. As imagens das diferenças de probabilidade de ocorrência estimadas (Figuras 8, 11, 14, 17 – Imagens “B”) identificam a similaridade entre os modelos. Com exceção de *Leporinus obtusidens*, as diferenças de probabilidade de ocorrência estimadas entre os modelos LOGIT e LOGIT_m são extremamente pequenas, não sendo representativas para estudos nesta escala. *Leporinus obtusidens* apresentou diferença entre os modelos na parte baixa da bacia, onde o modelo LOGIT informou trechos de montante como de ausência para a espécie e o modelo LOGIT_m informou estes mesmos trechos como de presença.

As diferenças entre os modelos LOGIT e LP, entretanto, apresentaram-se mais marcantes, seguindo um padrão similar para a todas as espécies. As maiores diferenças entre os dois modelos manifestaram-se nos segmentos de montante de rios de baixa altitude (Figuras 9, 12, 15, 18 – Imagens “B”), onde o modelo LOGIT estimou presenças mais à montante que o modelo LP.

Tabela 3: Coeficientes dos modelos de distribuição ajustados para *Salminus brasiliensis* na bacia do rio Uruguai segundo os modelos testados (veja o texto para mais detalhes). b_0 , b_1 e b_2 são coeficientes de calibração do modelo; *PMF* é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente; *TAXA* é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente.

		Média	Erro Padrão	Coef. de Variação	Wald	Significância
LOGIT	b_0	4,25590	1,69210	0,39759	6,32601	0,0119
	b_1	-1,95545	0,46091	0,23570	17,99967	<0,0001
	b_2	1,10370	0,22606	0,20482	23,83716	<0,0001
LOGIT _m	PMF _{Altitude (ln(m))}	5,4439	0,0853	0,0157	4072,4845	<0,0001
	PMF _{Área de Bacia (ln(km²))}	5,7678	0,0503	0,0087	13138,4266	<0,0001
	Taxa _{Altitude}	-1,8845	0,3240	0,1719	33,8380	<0,0001
	Taxa _{Área de Bacia}	1,0528	0,1583	0,1503	44,2490	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	231				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	320				
LP	PMF _{Altitude (ln(m))}	6,5097	0,0843	0,0129	77,2623	<0,0001
	PMF _{Área de Bacia (ln(km²))}	5,1148	0,3090	0,0604	16,5530	<0,0001
	Taxa _{Altitude}	4,2122	0,9546	0,2266	4,4128	<0,0001
	Taxa _{Área de Bacia}	1,0681	0,2285	0,2139	4,6741	<0,0001
	b_0	0,9935	0,0315	0,0318	31,4916	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	672				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	166				

Tabela 4: Coeficientes dos modelos de distribuição ajustados para *Prochilodus lineatus* na bacia do rio Uruguai segundo os modelos testados (veja o texto para mais detalhes). b_0 , b_1 e b_2 são coeficientes de calibração do modelo; *PMF* é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente; *TAXA* é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente.

		Média	Erro Padrão	Coef. de Variação	Wald	Significância
LOGIT	b_0	8,9884	4,0820	0,4541	4,8486	0,0277
	b_1	-2,6371	0,8901	0,3375	8,7768	0,0031
	b_2	0,9137	0,2579	0,2822	12,5527	0,0004
LOGIT _m	PMF_{Altitude} (ln(m))	5,0174	0,1755	0,0350	817,7160	<0,0001
	PMF_{Área de Bacia} (ln(km²))	4,6115	0,2284	0,0495	407,5636	<0,0001
	b_1	-4,5830	17,2893	3,7725	0,0703	0,7910
	b_2	1,6685	7,0226	4,2089	0,0564	0,8122
	$e^{(PMF_{Altitude})}$ (m)	151				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	101				
LP	PMF_{Altitude} (ln(m))	6,2458	0,0896	0,0143	69,7117	<0,0001
	PMF_{Área de Bacia} (ln(km²))	4,6189	0,5796	0,1255	7,9686	<0,0001
	Taxa_{Altitude}	3,9895	1,2359	0,3098	3,2281	<0,0001
	Taxa_{Área de Bacia}	0,6841	0,1682	0,2459	4,0664	<0,0001
	b_0	0,9971	0,0193	0,0193	51,6959	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	516				
$e^{(PMF_{Área de Bacia})}$ (km ²)	101					

Tabela 5: Coeficientes dos modelos de distribuição ajustados para *Leporinus obtusidens* na bacia do rio Uruguai segundo os modelos testados (veja o texto para mais detalhes). b_0 , b_1 e b_2 são coeficientes de calibração do modelo; *PMF* é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente; *TAXA* é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente.

		Média	Erro Padrão	Coef. de Variação	Wald	Significância
LOGIT	b_0	4,1290	1,9908	0,4821	4,3018	0,0381
	b_1	-1,9418	0,5288	0,2723	13,4838	0,0002
	b_2	1,1057	0,2240	0,2026	24,3549	<0,0001
LOGIT _m	PMF_{Altitude} (ln(m))	5,4423	0,1330	0,0244	1673,8565	<0,0001
	PMF_{Área de Bacia} (ln(km²))	5,1749	0,1827	0,0353	802,1760	<0,0001
	b_1	-2,4863	0,8463	0,3404	8,6307	0,0033
	b_2	0,9992	0,2086	0,2087	22,9552	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	231				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	177				
		PMF_{Altitude} (ln(m))	6,4780	0,0759	0,0117	85,3561
	PMF_{Área de Bacia} (ln(km²))	4,4214	0,4495	0,1017	9,8369	<0,0001
LP	Taxa_{Altitude}	4,4812	0,9007	0,2010	4,9753	<0,0001
	Taxa_{Área de Bacia}	0,7444	0,1476	0,1983	5,0425	<0,0001
	b_0	0,9997	0,0026	0,0026	379,8187	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	651				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	83				

Tabela 6: Coeficientes dos modelos de distribuição ajustados para *Pseudoplatystoma corruscans* na bacia do rio Uruguai segundo os modelos testados (veja o texto para mais detalhes). b_0 , b_1 e b_2 são coeficientes de calibração do modelo; *PMF* é o ponto de mudança de fase de cada parâmetro, indicando a altitude ou área de bacia que redundariam em uma probabilidade de ocorrência de 50%, caso o parâmetro fosse analisado isoladamente; *TAXA* é um parâmetro indicativo da velocidade relativa de mudança de estado ausente/presente.

		Média	Erro Padrão	Coef. de Variação	Wald	Significância
LOGIT	b_0	-2,7198	2,5079	0,9221	1,1761	0,2781
	b_1	-1,5875	0,3635	0,2290	19,0731	0,0000
	b_2	1,4244	0,4454	0,3127	10,2290	0,0014
LOGIT _m	PMF_{Altitude} (ln(m))	4,2236	0,1949	0,0462	469,5032	<0,0001
	PMF_{Área de Bacia} (ln(km²))	6,5642	0,2593	0,0395	640,7284	<0,0001
	b_1	-1,6149	0,3786	0,2344	18,1963	<0,0001
	b_2	1,3815	0,3304	0,2392	17,4842	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	68				<0,0001
	$e^{(PMF_{Área de Bacia})}$ (km ²)	709				<0,0001
LP	PMF_{Altitude} (ln(m))	6,2332	0,1492	0,0239	41,7706	<0,0001
	PMF_{Área de Bacia} (ln(km²))	7,3921	0,2501	0,0338	29,5554	<0,0001
	Taxa_{Altitude}	3,9407	1,9237	0,4881	2,0486	0,0203
	Taxa_{Área de Bacia}	1,5830	1,0480	0,6620	1,5105	0,0655
	b_0	0,9927	0,0122	0,123	81,3824	<0,0001
	$e^{(PMF_{Altitude})}$ (m)	509				
	$e^{(PMF_{Área de Bacia})}$ (km ²)	1623				

Tabela 7: Critérios estatísticos para seleção de modelos executados com os modelos LOGIT, LOGIT_m e LP. *AIC_m* = Akaike Information Criterion (conforme modificação descrita pela equação 5); *DSI* = Dimensional Stability Index. *Aderência Total* representa o percentual de acerto do modelo, *Aderência Positiva* representa o percentual de acerto do modelo para presença e *Aderência Negativa* representa o percentual de acerto do modelo para ausência.

	Modelo	Aderência			Critério Estatístico		
		Total	Positiva	Negativa	AIC	DSI	Variância Residual
<i>Salminus brasiliensis</i>	LOGIT	81,44	84,16	77,27	27,1578	0,2677	0,1291
	LOGIT _m	81,44	84,16	77,27	29,1578	0,0609	0,1298
	LP	77,25	84,16	66,66	30,4726	0,0655	0,1379
<i>Prochilodus lineatus</i>	LOGIT	82,04	82,02	83,12	27,0396	0,3511	0,1283
	LOGIT _m	82,04	83,15	81,82	30,0664	0,4072	0,1354
	LP	79,64	78,65	81,82	31,2810	0,0767	0,1428
<i>Leposinus obtusidens</i>	LOGIT	81,44	83,17	78,79	27,1739	0,2985	0,1291
	LOGIT _m	77,25	84,62	65,08	30,0244	0,0885	0,1351
	LP	79,42	86,54	66,66	33,3084	0,0416	0,1438
<i>Pseudoplatystoma corruscans</i>	LOGIT	84,43	67,31	92,17	23,5041	0,4042	0,1067
	LOGIT _m	85,03	67,31	93,04	25,4904	0,1006	0,1073
	LP	83,23	67,31	90,44	26,6114	0,0797	0,1124

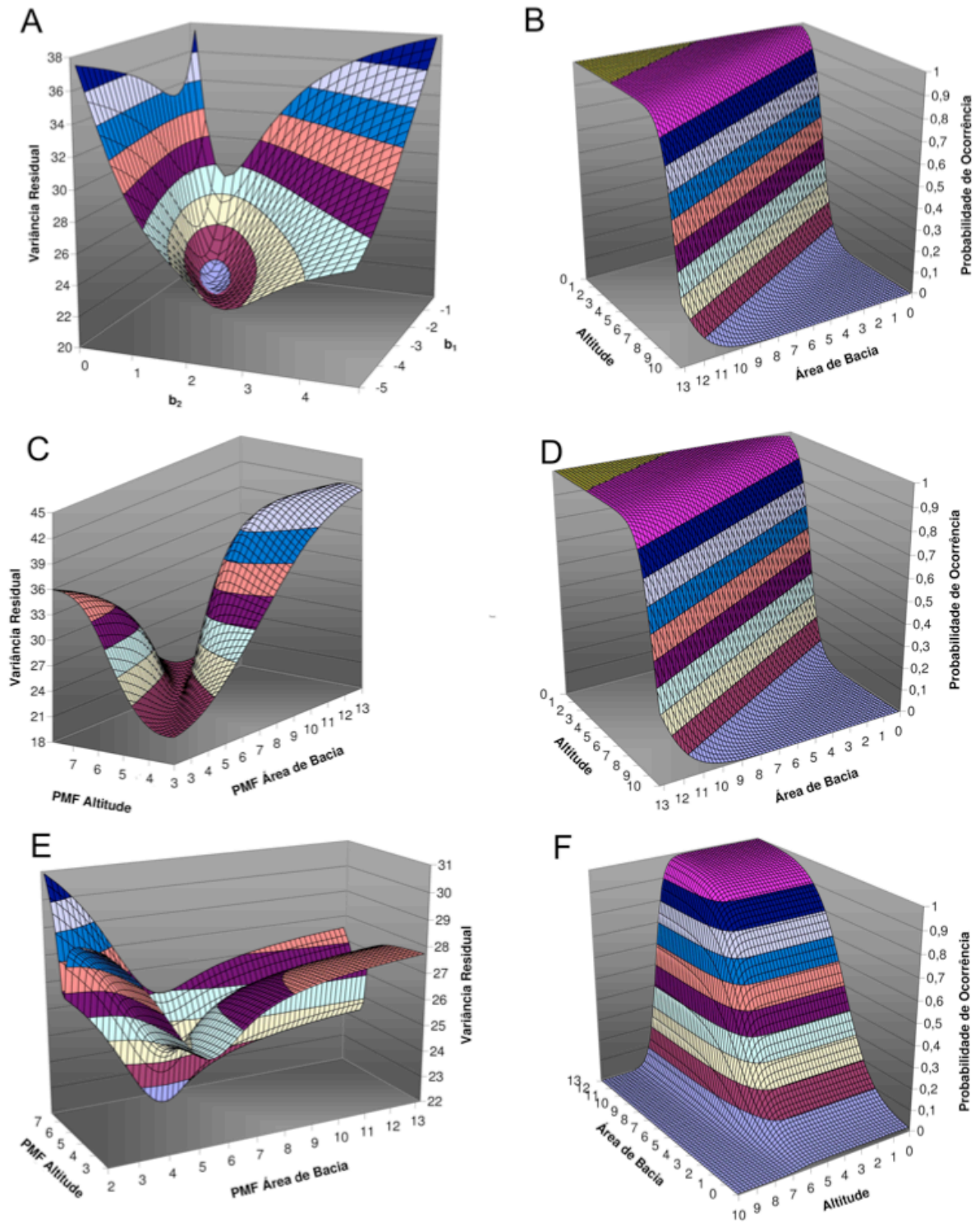


Figura 3: Campo dimensional formado pela Soma do Quadrado dos Resíduos (SQR) e probabilidade de ocorrência obtidos com os modelos *LOGIT* (“A” e “B”), *LOGIT_m* (“C” e “D”) e *LP* (“E” e “F”) para *Salminus brasiliensis* na bacia Uruguai. O modelo *LOGIT* não apresenta os parâmetros Ponto de Mudança de Fase, dessa forma foram utilizados os parâmetros b_1 e b_2 como variáveis independentes na construção da imagem “A”. As variáveis independentes dos demais modelos são representadas em escala logarítmica, $\ln(\text{km}^2)$ para variáveis referentes a área de bacía e $\ln(\text{m})$ para variáveis referentes a altitude.

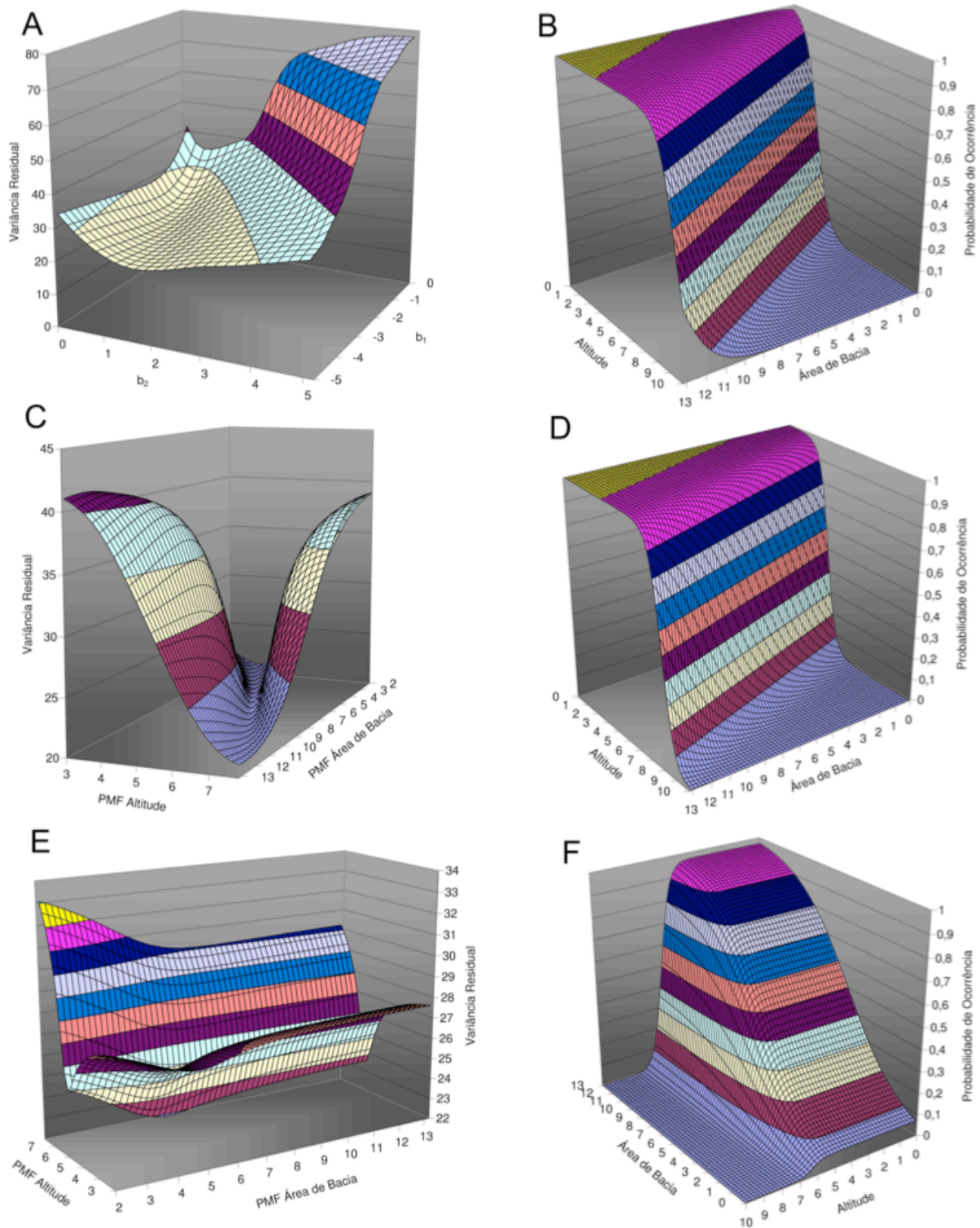


Figura 4: Campo dimensional formado pela Soma do Quadrado dos Resíduos (SQR) e probabilidade de ocorrência obtidos com os modelos *LOGIT* (“A” e “B”), *LOGIT_m* (“C” e “D”) e *LP* (“E” e “F”) para *Prochilodus lineatus* na bacia Uruguai. O modelo *LOGIT* não apresenta os parâmetros Ponto de Mudança de Fase, dessa forma foram utilizados os parâmetros b_1 e b_2 como variáveis independentes na construção da imagem “A”. As variáveis independentes dos demais modelos são representadas em escala logarítmica, $\ln(\text{km}^2)$ para variáveis referentes a área de bacia e $\ln(\text{m})$ para variáveis referentes a altitude.

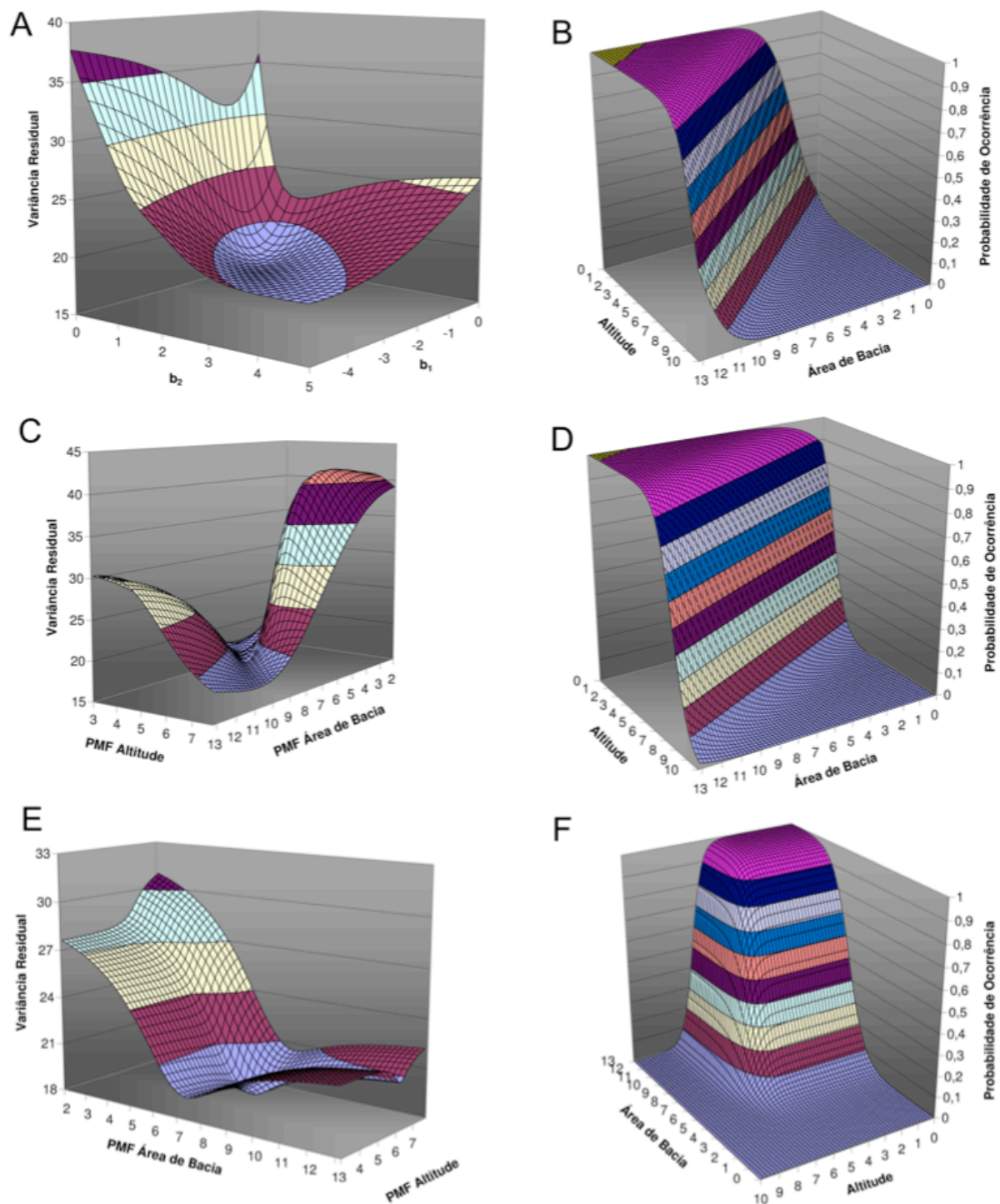


Figura 5: Campo dimensional formado pela Soma do Quadrado dos Resíduos (SQR) e probabilidade de ocorrência obtidos com os modelos LOGIT (“A” e “B”), LOGIT_m (“C” e “D”) e LP (“E” e “F”) para *Pseudoplatystoma corruscans* na bacia Uruguai. O modelo LOGIT não apresenta os parâmetros Ponto de Mudança de Fase, dessa forma foram utilizados os parâmetros b_1 e b_2 como variáveis independentes na construção da imagem “A”. As variáveis independentes dos demais modelos são representadas em escala logarítmica, $\ln(\text{km}^2)$ para variáveis referentes a área de bacia e $\ln(\text{m})$ para variáveis referentes a altitude.

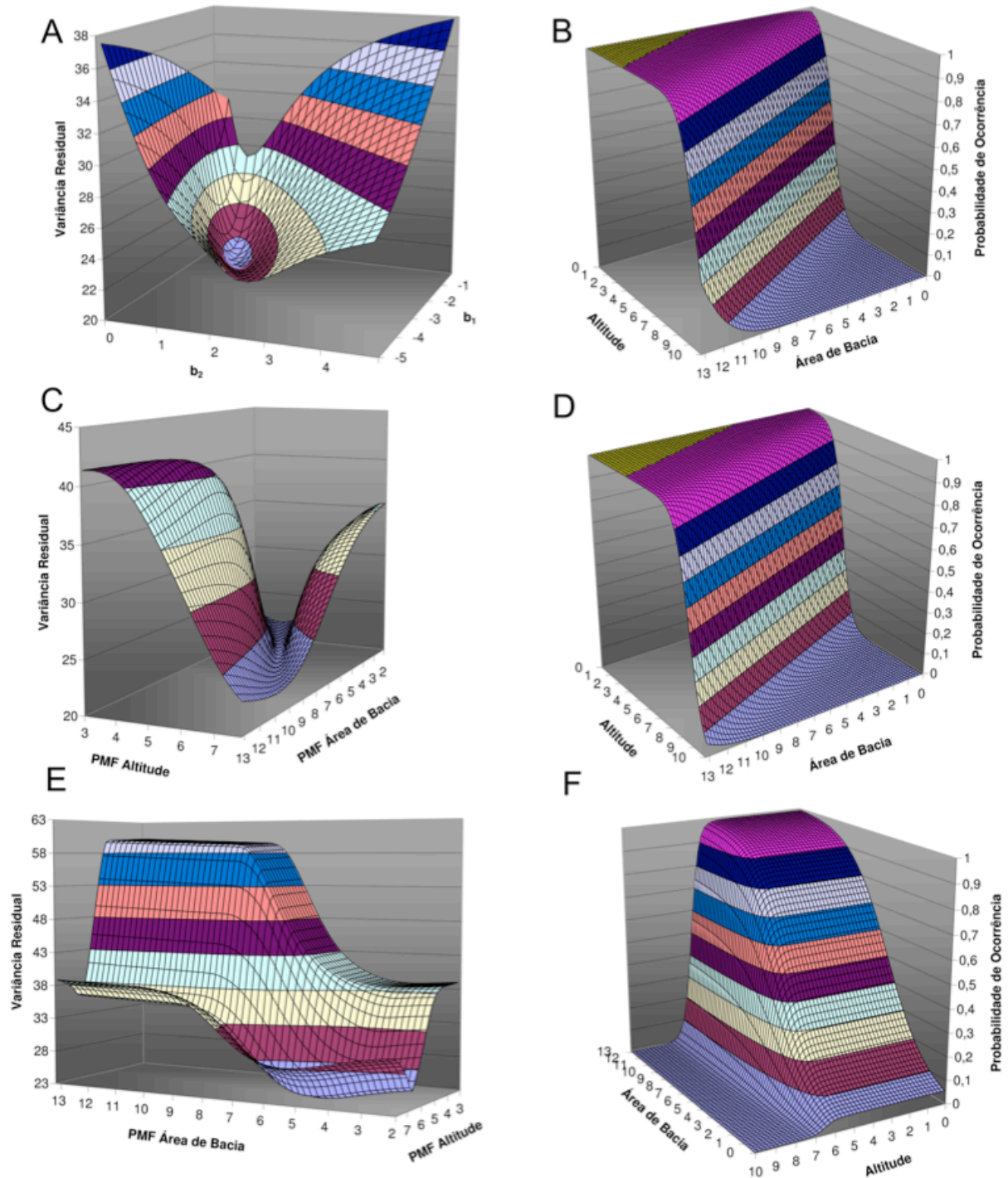


Figura 6: Campo dimensional formado pela Soma do Quadrado dos Resíduos (SQR) e probabilidade de ocorrência obtidos com os modelos *LOGIT* (“A” e “B”), *LOGIT_m* (“C” e “D”) e *LP* (“E” e “F”) para *Leporinus obtusidens* na bacia Uruguai. O modelo *LOGIT* não apresenta os parâmetros Ponto de Mudança de Fase, dessa forma foram utilizados os parâmetros b_1 e b_2 como variáveis independentes na construção da imagem “A”. As variáveis independentes dos demais modelos são representadas em escala logarítmica, $\ln(\text{km}^2)$ para variáveis referentes a área de bacía e $\ln(\text{m})$ para variáveis referentes a altitude.

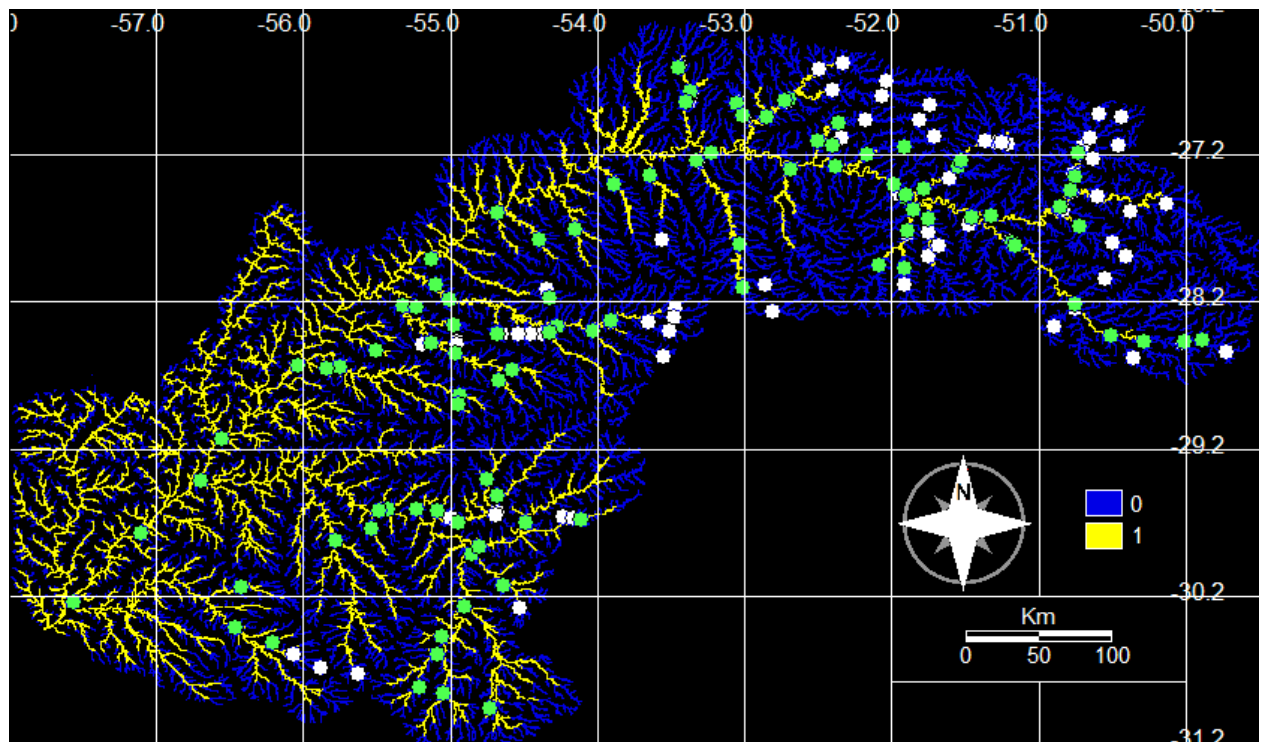


Figura 7: *Salminus brasiliensis* – Mapa de distribuição gerado pelo modelo LOGIT para toda a bacia hidrográfica do rio Uruguai. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

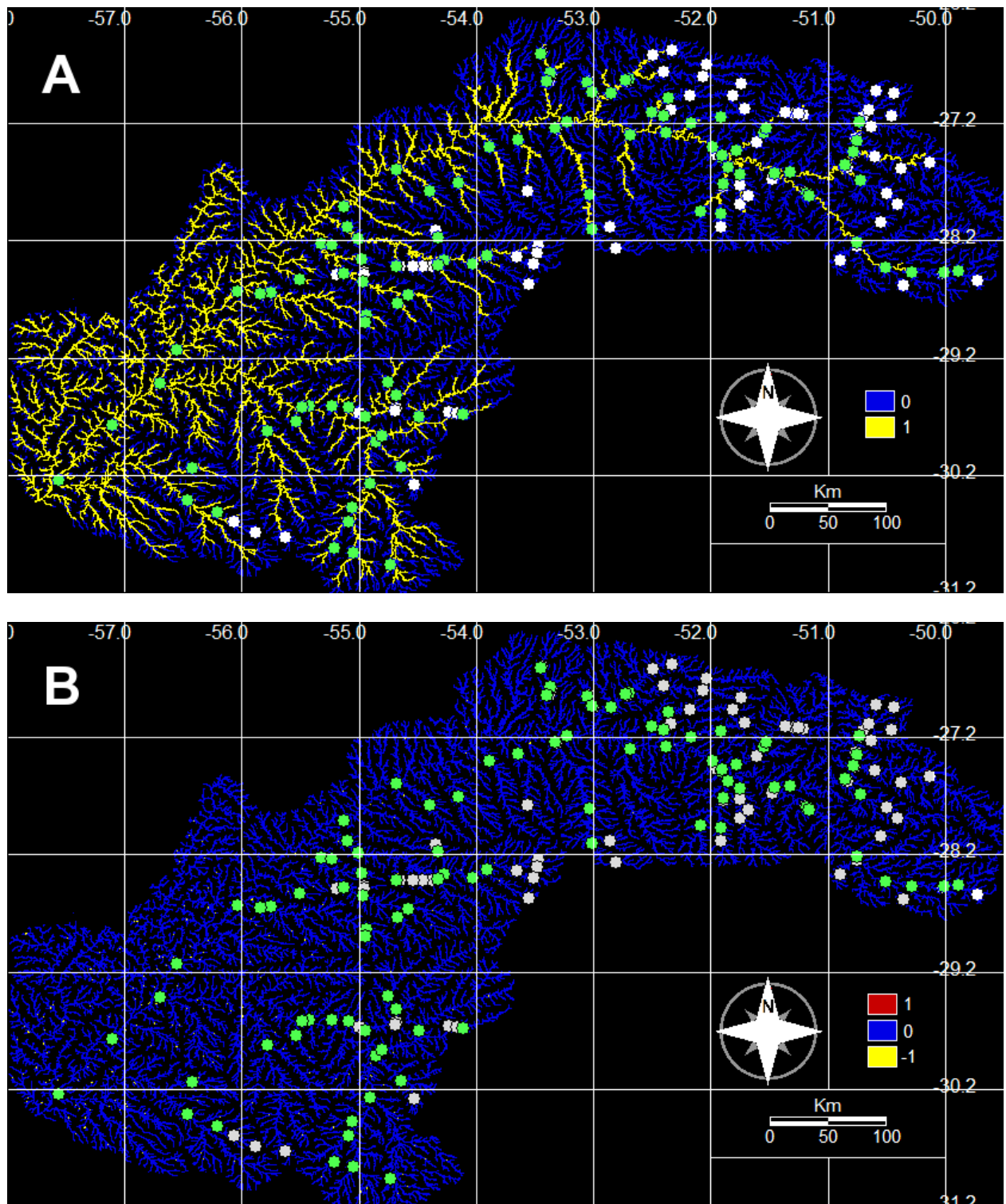


Figura 8: *Salminus brasiliensis* – A) Mapa de distribuição gerado pelo modelo LOGIT_m para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LOGIT_m. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LOGIT_m, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LOGIT_m. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

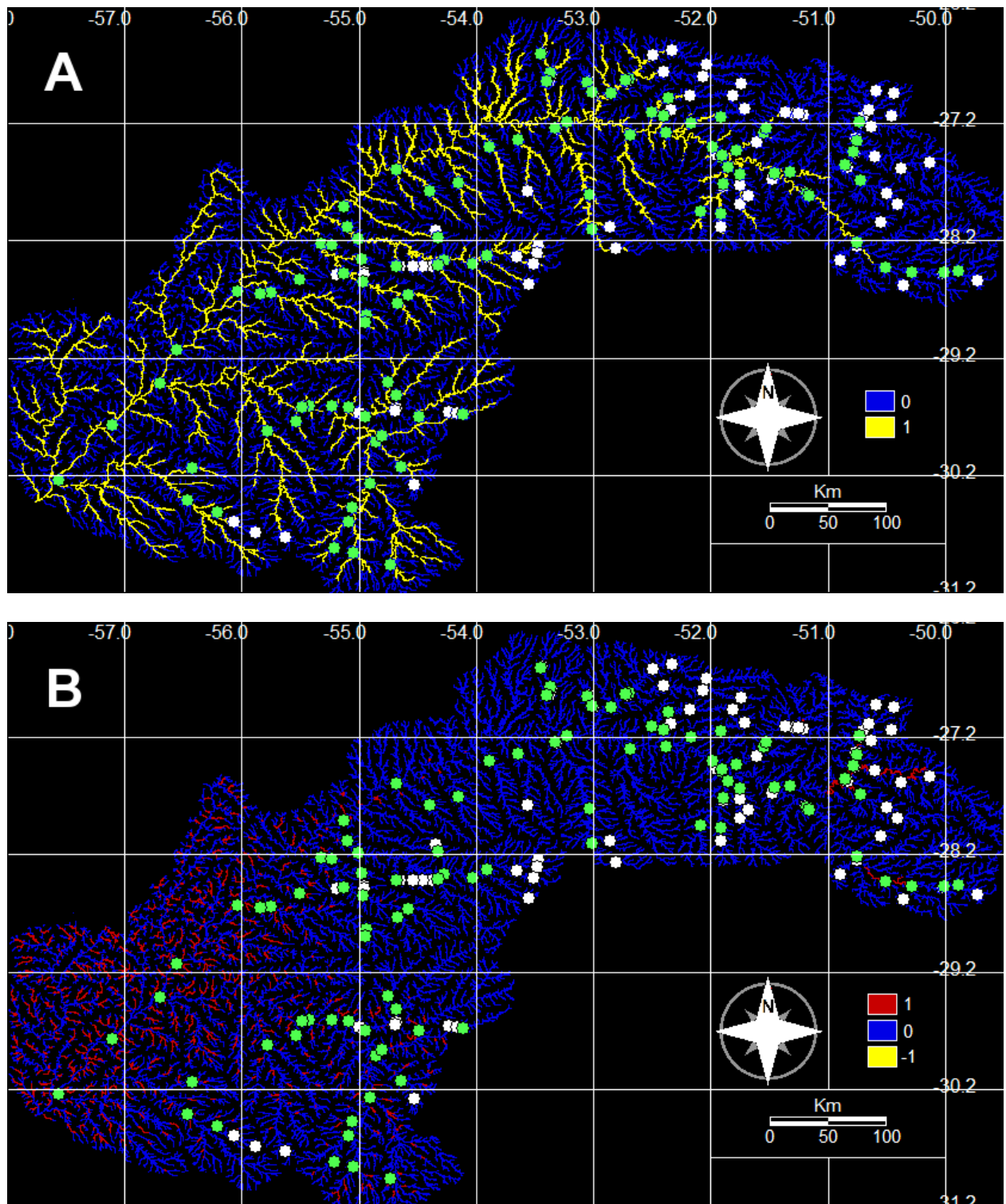


Figura 9: *Salminus brasiliensis* – A) Mapa de Distribuição gerado pelo modelo LP para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LP. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LP, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LP. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

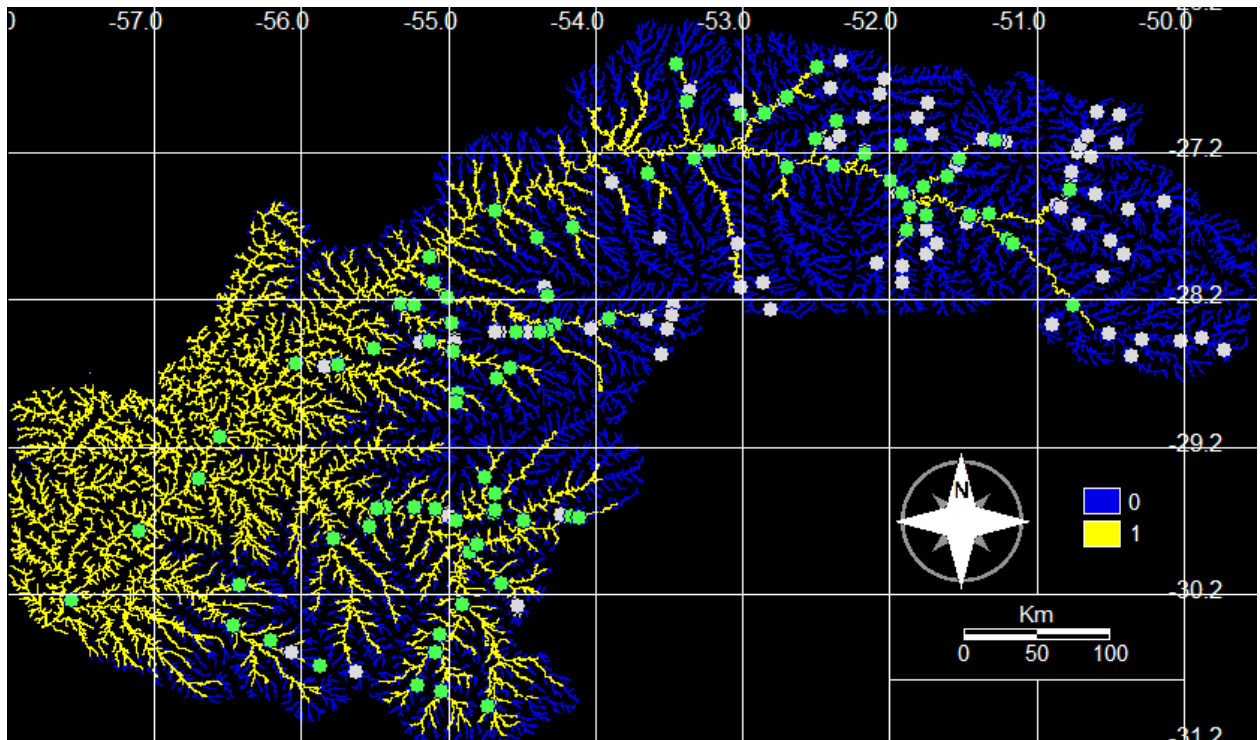


Figura 10: *Prochilodus lineatus* – Mapa de distribuição gerado pelo modelo LOGIT para toda a bacia hidrográfica do rio Uruguai. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência obtidas.

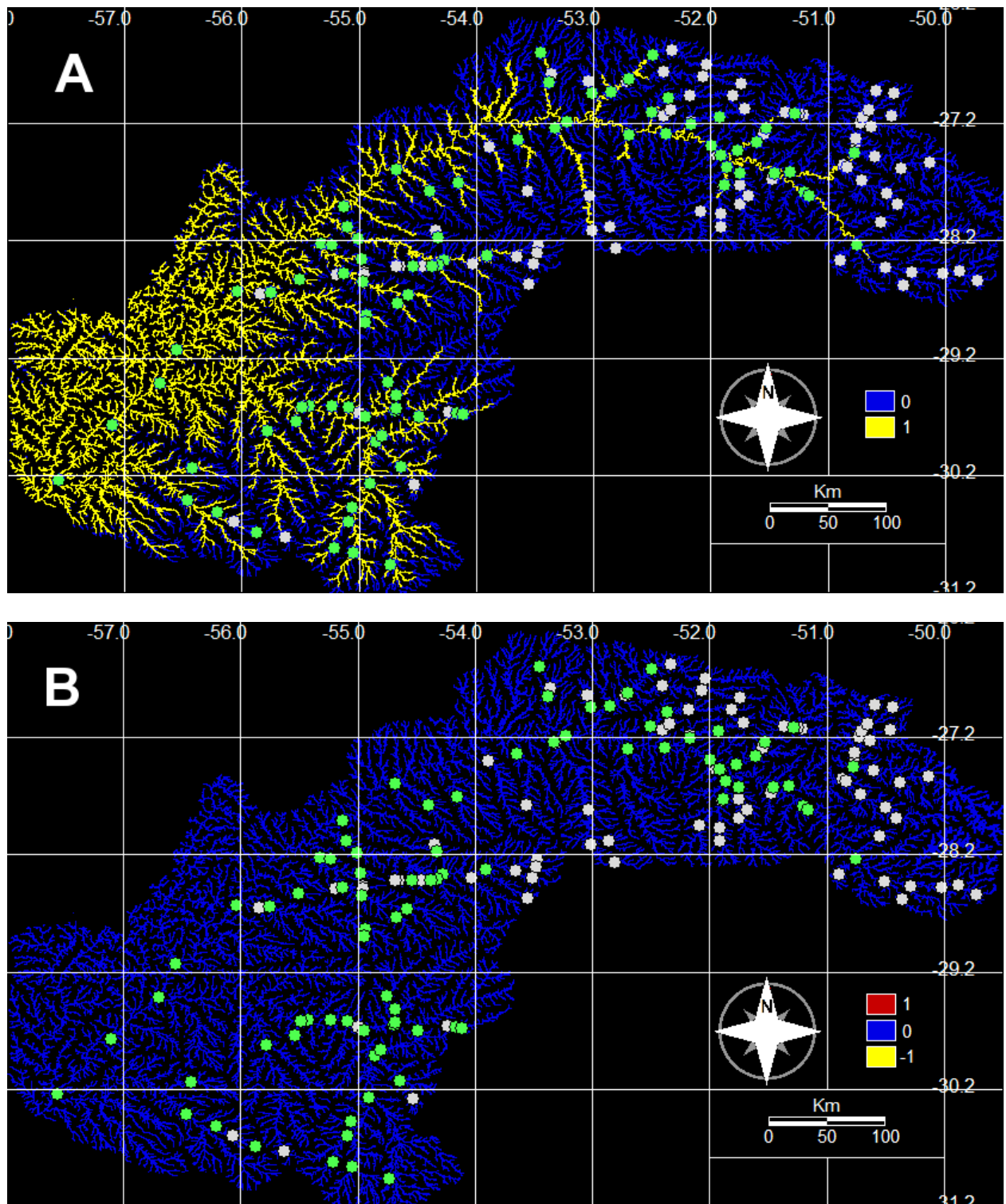


Figura 11: *Prochilodus lineatus* – A) Mapa de distribuição gerado pelo modelo LOGIT_m para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LOGIT_m. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LOGIT_m, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LOGIT_m. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

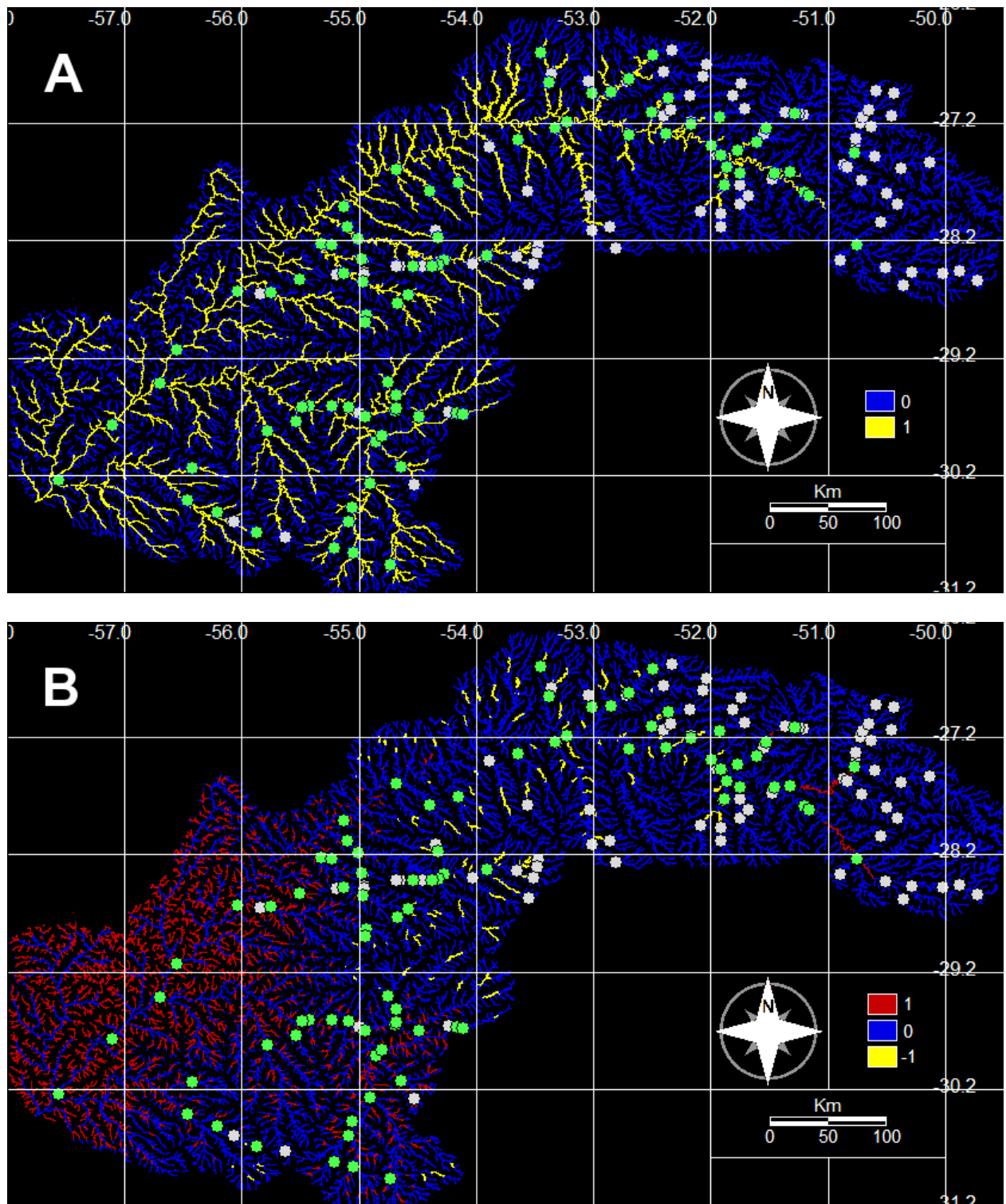


Figura 12: *Prochilodus lineatus* – A) Mapa de Distribuição gerado pelo modelo LP para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LP. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LP, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LP. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

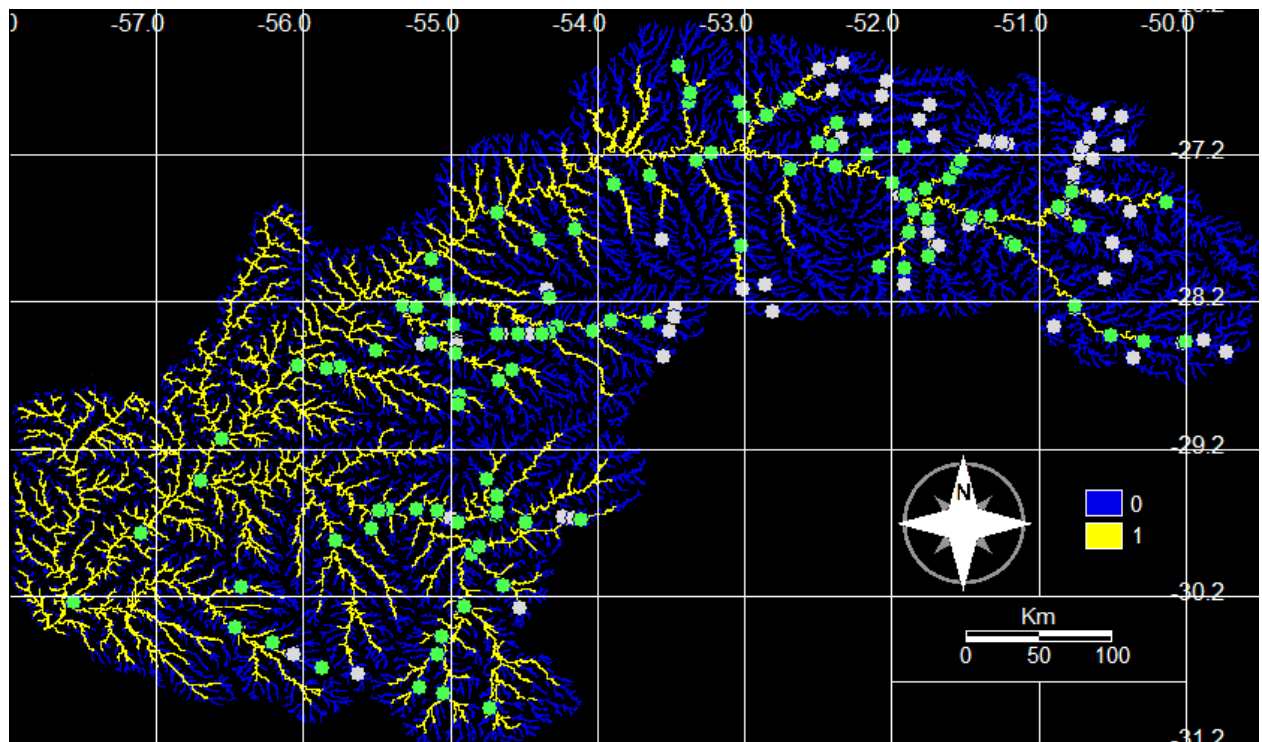


Figura 13: *Leporinus obtusidens* – Mapa de Distribuição gerado pelo modelo LOGIT para toda a bacia hidrográfica do rio Uruguai. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

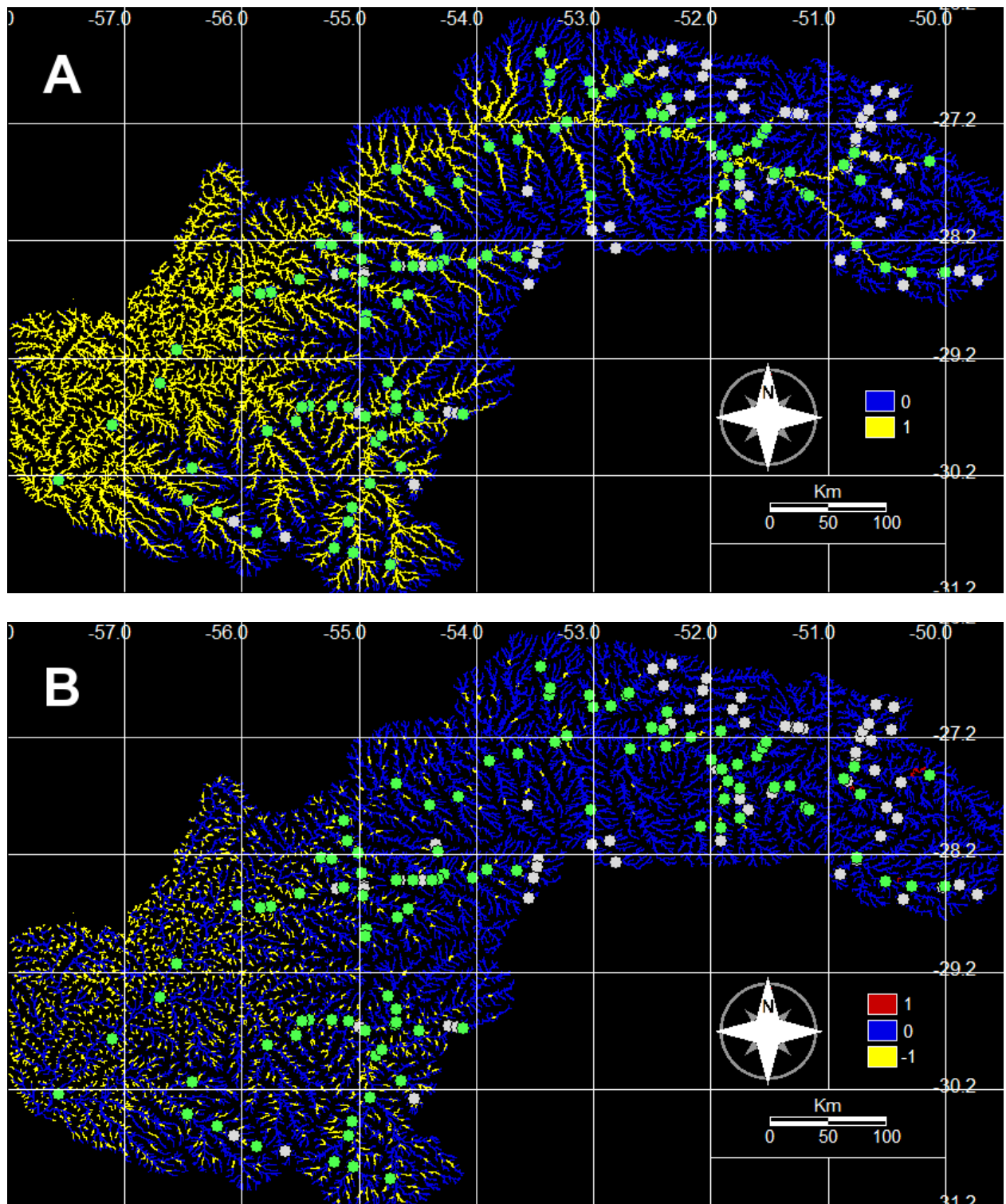


Figura 14: *Leporinus obtusidens* – A) Mapa de Distribuição gerado pelo modelo LOGIT_m para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LOGIT_m. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LOGIT_m, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LOGIT_m. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência, obtidas por entrevistas.

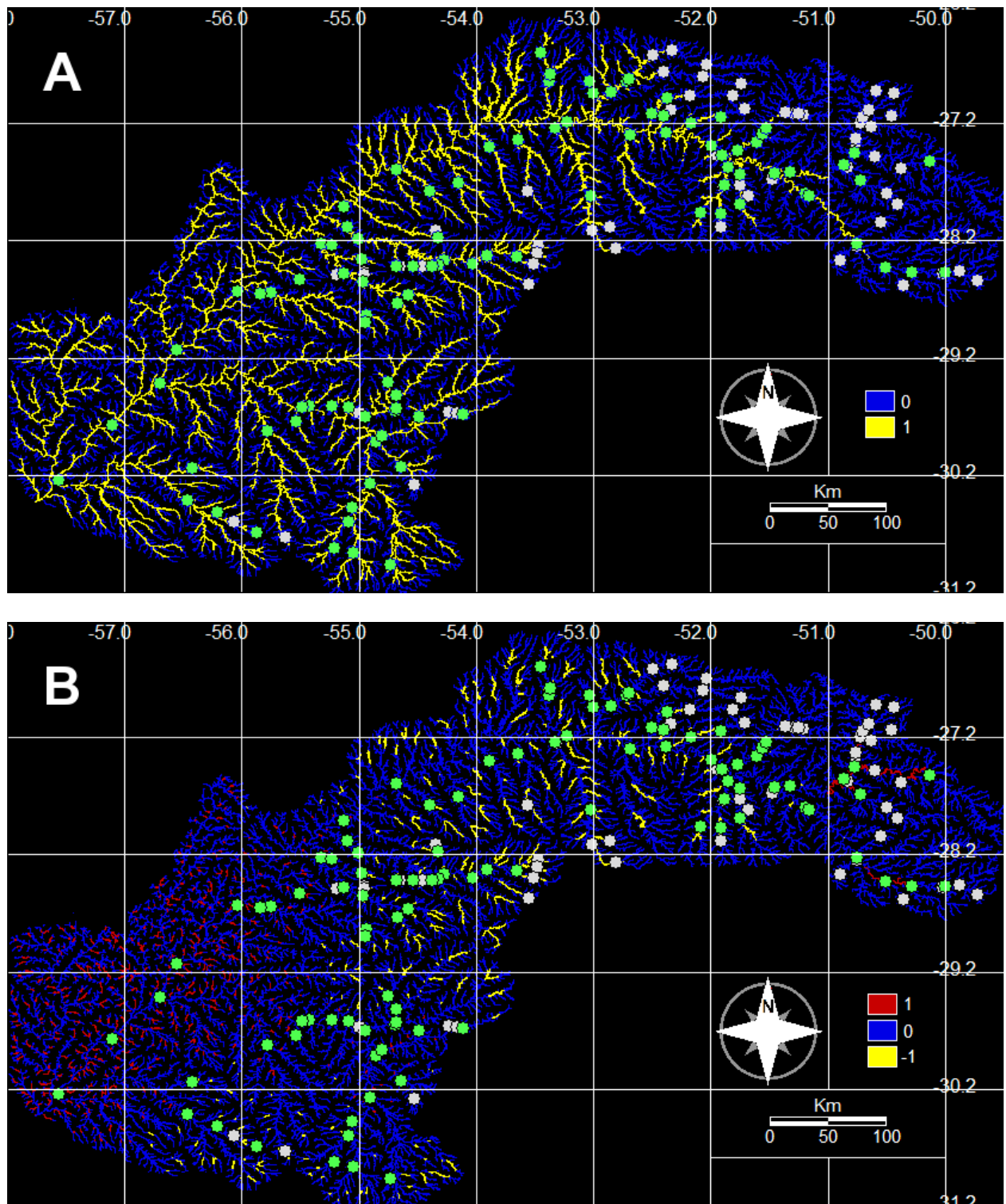


Figura 15: *Leporinus obtusidens* – A) Mapa de Distribuição gerado pelo modelo LP para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LP. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LP, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LP. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

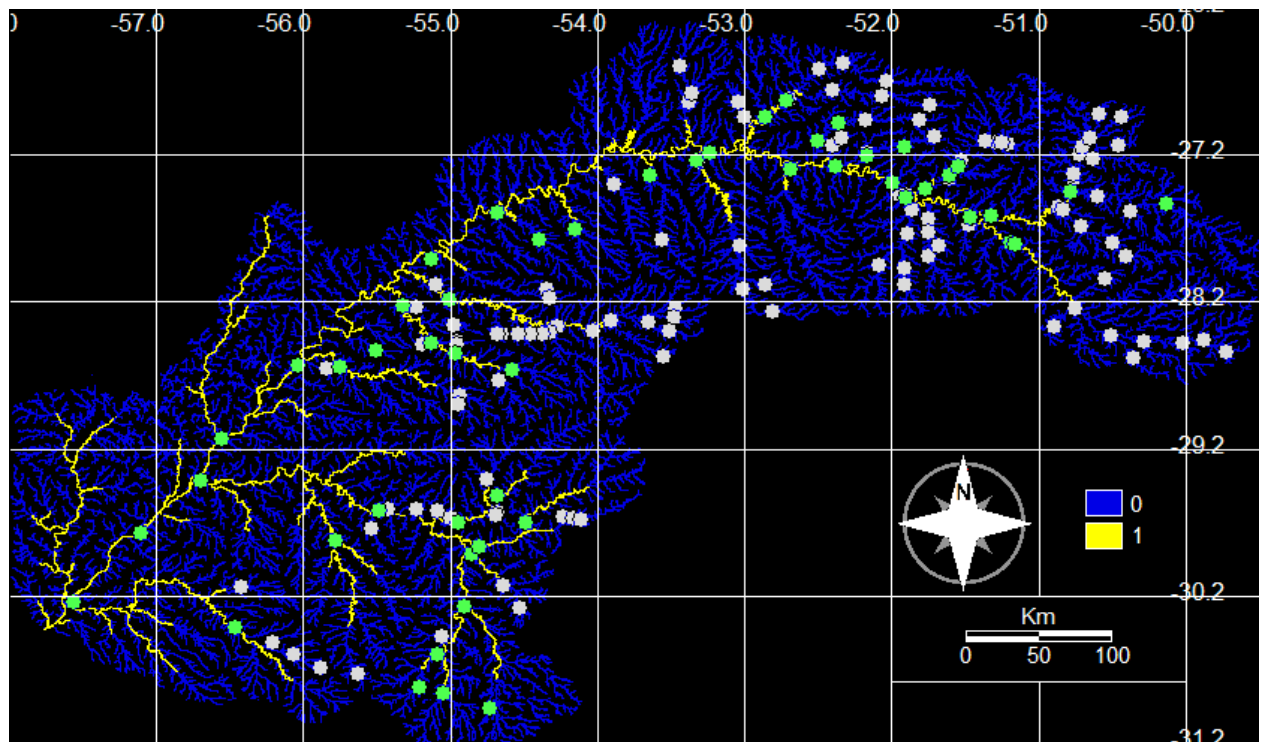


Figura 16: *Pseudoplatysoma corruscans* – Mapa de Distribuição gerado pelo modelo LOGIT para toda a bacia hidrográfica do rio Uruguai. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

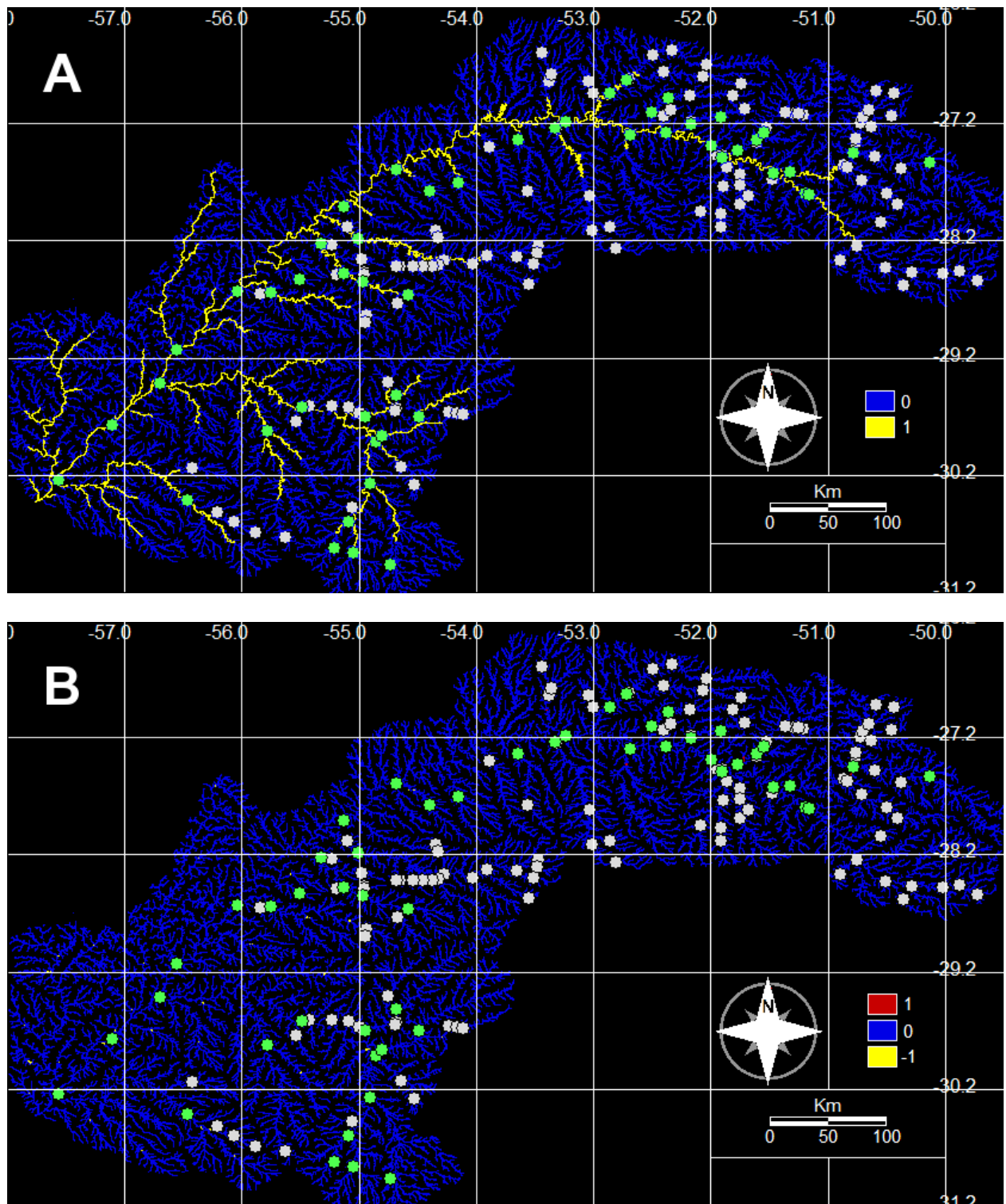


Figura 17: *Pseudoplatystoma corruscans* – A) Mapa de Distribuição gerado pelo modelo LOGIT_m para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LOGIT_m. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LOGIT_m, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LOGIT_m. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

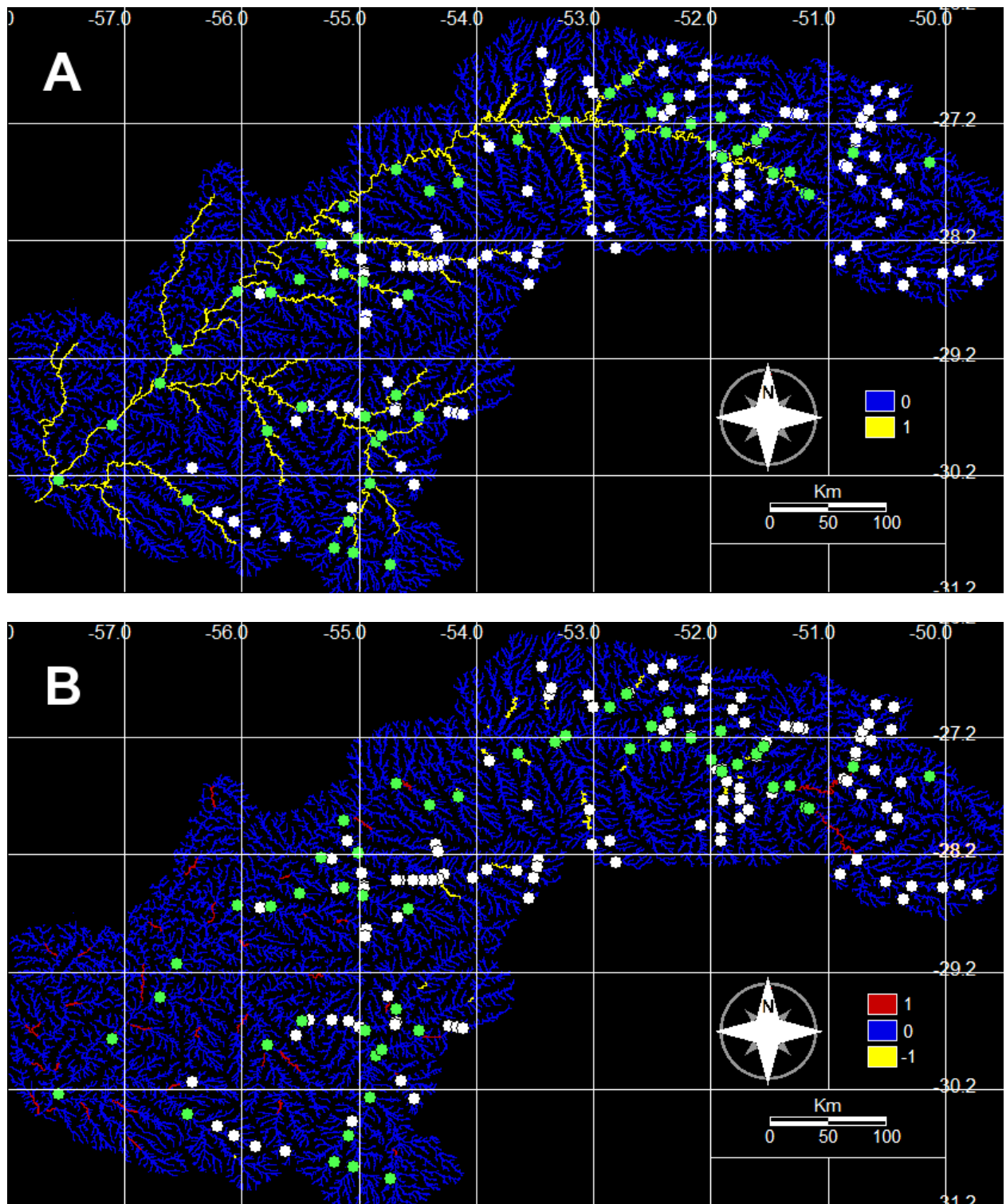


Figura 18: *Pseudoplatystoma corruscans* – A) Mapa de Distribuição gerado pelo modelo LP para toda a bacia hidrográfica do rio Uruguai. B) Mapa de diferença entre os modelos LOGIT e LP. Trechos em vermelho informam locais de presença para o modelo LOGIT e ausência para o modelo LP, trechos em amarelo informam locais de ausência para o modelo LOGIT e presença para o modelo LP. Pontos em verde e branco representam, respectivamente, as informações de presença e ausência.

4. DISCUSSÃO

A natureza é muito complexa e heterogênea para ser prevista com precisão em todos os aspectos de tempo e espaço a partir de um único modelo (GUISAN & ZIMMERMANN, 2000). Entretanto, avanços nos campos de ciência da computação, estatística e a necessidade urgente de se conhecer a distribuição das espécies tem tornado seu uso cada vez mais frequente (PEARCE & FERRIER, 2000a; GUIBAN & THUILLER, 2005; GONÇALVES, 2010). O geoprocessamento é uma importante ferramenta para gestão ambiental, visto que pode combinar informações referentes às exigências ambientais de espécies com modelos numéricos de terreno derivados de imagens de satélites, produzindo, assim, mapeamentos temáticos, diagnósticos ambientais, avaliações de impactos ambiental, ordenamentos territoriais e prognósticos ambientais diversos (MEDEIROS & CÂMARA, 2001; TURNER *et al.*, 2003; GONÇALVES, 2010).

A proposta de um novo modelo neste trabalho tem o objetivo apresentar uma nova possibilidade para modelar a distribuição de espécies. Alves & Fontoura (2009) e Gonçalves (2010) demonstraram que o modelo LOGIT é eficiente para modelar a distribuição de peixes em diferentes bacias. Entretanto, quando testado em situações-limite, o modelo LOGIT pode apresentar-se limitado com relação à sua capacidade preditiva de extrapolação para condições não contempladas na matriz de dados.

O ajuste de um modelo depende da escolha de medidas que representem a adequação do modelo à matriz de dados e minimizem as diferenças entre os valores observados e obtidos através da modelagem (FRANKLIN, 2010). Ao ajustar um modelo através da soma do quadrado dos resíduos, busca-se sempre a condição de mínima variância residual, ou seja, o ponto onde o resultado alcançado pelo modelo é o mais parecido com os dados reais. O problema encontrado no modelo LOGIT_m está nessa questão: quando calibrado utilizando apenas duas variáveis descritoras, inúmeras soluções de mesma qualidade estatística foram alcançadas. Este comportamento ocorre devido ao desenho do modelo. O modelo LOGIT_m trabalha com blocos relativos a cada variável independente, onde temos o produto de um coeficiente de ajuste (b_1, b_2, \dots, b_i) pela diferença de cada variável em relação ao seu ponto de mudança de fase ($x_1 - PMF_1, x_2 - PMF_2, \dots, x_i - PMF_i$), sendo cada bloco ($b_i \cdot [x_i - PMF_i]$) ligado ao seguinte através de uma soma ($b_1 \cdot [x_1 - PMF_1] + b_2 \cdot [x_2 - PMF_2] + \dots + b_i \cdot [x_i - PMF_i]$). Os coeficientes b ponderam o efeito relativo da variável proporcionalmente à sua unidade de medida. Entretanto, durante o processo de ajuste estes parâmetros podem agir de maneira compensatória: mesmo que uma variável

determine uma baixa probabilidade de ocorrência, como, por exemplo, uma área de bacia muito baixa, seu efeito pode ser mascarado compensatoriamente por outra variável, como uma altitude muito baixa, tornando possível que sejam alcançados resultados irreais.

O modelo LOGIT consegue, de certa forma, corrigir parte do erro causado durante o ajuste devido à soma direta da variável b_0 aos blocos de variáveis independentes, utilizando este parâmetro como um fator de correção (Figuras 3 – 6, imagens “A”). Entretanto, como cada bloco é ligado ao seguinte através de uma soma e as variáveis de altitude e área de bacia são inversamente proporcionais, a condição compensatória do modelo permanece após a calibração do mesmo. As figuras 3 – 6, imagens “B”, evidenciam este processo, pois o modelo considera as espécies como presentes, com probabilidades próximas a 100%, em locais com valores de área de bacia extremamente baixos, bastando para isso que tais pontos tenham pequenos valores de altitude, apresentando viés similar ao modelo LOGIT_m.

A modificação proposta por Gonçalves (2010) não altera o *design* básico do modelo LOGIT, embora exista a adição dos parâmetros Ponto de Mudança de Fase. A modificação consistiu em desmembrar a constante b_0 , que está atrelada igualmente a todas as variáveis, em coeficientes relativos a cada variável ambiental. A sequência de equações (Equações 9 – 12) a seguir mostra a alteração feita por Gonçalves (2010).

$$\text{LOGIT}_m = P = e^{(b_1(x_1 + \text{PMF}_1) + b_2(x_2 + \text{PMF}_2))} \cdot (1 + e^{(b_1(x_1 + \text{PMF}_1) + b_2(x_2 + \text{PMF}_2))})^{-1} \quad (7)$$

logo,

$$P = e^{(b_1 x_1 + b_1 \text{PMF}_1 + b_2 x_2 + b_2 \text{PMF}_2)} \cdot (1 + e^{(b_1 x_1 + b_1 \text{PMF}_1 + b_2 x_2 + b_2 \text{PMF}_2)})^{-1} \quad (8)$$

$$b_0 = b_1 \text{PMF}_1 + b_2 \text{PMF}_2 \quad (9)$$

$$\text{LOGIT} = P = e^{(b_0 + b_1 x_1 + b_2 x_2)} \cdot (1 + e^{(b_0 + b_1 x_1 + b_2 x_2)})^{-1} \quad (10)$$

Sendo mantida, então, a forma do modelo LOGIT, a interação entre os parâmetros ambientais continuou compensatória durante o cálculo de ocorrência das espécies. Além disso, ao desmembrar a constante b_0 e incluí-la nos blocos de variáveis ambientais, perdeu-se o comportamento de “fator de correção” durante a calibração do modelo, resultando em um comportamento de “calha” de pequenos valores de variância residual. Sendo assim, os coeficientes podem ser ajustados em condições não possíveis naturalmente, ocasionando em taxas de transição presença/ausência muito baixas ou muito altas, ou mesmo PMFs deslocados para

pontos nos extremos da bacia, o que resultaria em resultados de presença ou ausência não correspondentes à realidade.

O modelo LP resultou em apenas um ponto de melhor ajuste para todas as espécies. Em geral, as aderências alcançadas foram ligeiramente mais baixas no novo modelo. O menor percentual de aderência obtido com o modelo LP foi próximo a 67%. A redução na capacidade preditiva do modelo em alguns casos se deve aos valores de SQR levemente mais elevados, entretanto, o pequeno aumento no SQR é compensado pela capacidade do modelo de convergir sempre para o mesmo ponto em todo o campo dimensional.

Embora o modelo LOGIT tenha alcançado um padrão semelhante de convergência do campo dimensional de SQR, o mesmo apresenta problema com relação ao campo dimensional de probabilidade de ocorrência (Figuras 4 – 7, imagens “B”), possibilitando estimativas irreais em situações-limite. Tal problema foi solucionado no modelo LP, pois, conforme visto nas figuras 4 – 7, imagens “F”, as altas probabilidades de ocorrência concentram-se nas combinações de pequenas altitudes e grandes áreas de bacia.

O critério estatístico AIC (Equação 5) sofreu modificações para ser aplicado neste estudo. Originalmente o AIC trabalha utilizando o teste qui-quadrado como um de seus parâmetros, assim o teste é composto, em parte, pela penalização do modelo conforme sua adequabilidade ao problema estudado. A alteração na equação do AIC foi necessária porque o uso do teste qui-quadrado não se mostrou eficiente para o problema abordado neste projeto. Como os valores do teste qui-quadrado são obtidos através da divisão dos resíduos pelos valores alcançados pelo modelo, ocorreu um erro nos pontos onde a predição do modelo foi próxima de zero e a informação obtida em campo era de presença, pois a divisão por valores muito pequenos resultou em quocientes extremamente altos, não representando a real condição de adequabilidade do modelo. Para que fosse mantida a característica do método de avaliar a adequabilidade, optamos por trabalhar diretamente com o valor de SQR, pois este também indica diretamente a capacidade de ajuste do modelo. Assim, o formato do AIC foi mantido, onde o melhor modelo é aquele que obtém o valor mais próximo de zero.

O modelo LP alcançou valores de AIC mais altos que os modelos LOGIT e LOGIT_m para todas as espécies (Tabela 3). O LP é um modelo mais complexo que o LOGIT e o LOGIT_m, trabalhando com 5 parâmetros quando os demais utilizam três (LOGIT) e quatro (LOGIT_m). Teoricamente, um modelo complexo tem melhor capacidade de ajuste que um modelo mais simples, daí a importância de penalizar um

modelo conforme seu número de parâmetros. Contudo, esse resultado não foi encontrado no presente trabalho, onde o modelo LP alcançou valores de SQR ligeiramente mais altos que os demais. Somando estes dois fatores, maior número de parâmetros e SQR mais elevado, explica-se o resultado alcançado pelo critério AIC.

Tendo em vista que o grande problema para os modelos já descritos concentra-se no formato de seus campos dimensionais de SQR e probabilidade de ocorrência, desenvolvemos neste trabalho o método DSI (Equação 7). Este critério penaliza os modelos conforme o formato do campo de SQR. O coeficiente de variação de cada parâmetro está ligado diretamente a heterogeneidade do campo resultante de um modelo, quanto mais heterogêneo for o campo, menor será o coeficiente de variação dos parâmetros em função da máxima convergência para um resultado único. O DSI reflete, então, uma representação numérica do campo dimensional apresentado nas figuras 4 – 7, imagens “A”, “C” e “E”.

O modelo LP foi o que obteve, de forma geral, os melhores resultados no critério DSI (Tabela 3). Nesta análise, o modelo LOGIT_m apresentou melhores resultados que o modelo LOGIT, embora este último tenha mostrado comportamento melhor quanto a regularidade do formato do campo de SQR (Figuras 4 – 7, imagens “A” e “C”), e resultado ligeiramente melhor que o modelo LP quando aplicado para *Salminus brasiliensis*. A distorção deste resultado é devido a forma da “calha de menor variância” formada pelo modelo LOGIT_m: como o modelo age de forma compensatória entre as duas variáveis, os menores valores de SQR se dão diagonalmente no campo, estendendo-se das combinações de mínimas altitudes e áreas de bacia até as combinações de máximas altitude e áreas de bacia (Figuras 4 – 7, imagens “A” e “C”). O modelo de avaliação DSI trabalha paralelamente ao eixo de cada parâmetro, sofrendo interferência da grande variação apresentada no mesmo sentido dos eixos de PMF_{Altitude} e PMF_{Área de Bacia} do modelo LOGIT_m. Portanto, o índice gerado pelo critério DSI não pode ser analisado isoladamente: a observação dos campos dimensionais faz parte do critério de avaliação dos modelos e não pode ser descartada.

Os modelos foram analisados ainda quanto a sua variância residual (Tabela 3). LOGIT e LOGIT_m obtiveram resultados melhores que o modelo LP, entretanto pode-se dizer que os modelos comportaram-se de forma equivalente neste critério, pois a diferença entre os três modelos foi de pequena monta.

Um aspecto relevante deste estudo é a comparação de modelos através de produtos cartográficos da área estudada. As Figuras 8 – 19 apresentam os resultados obtidos com os três modelos para toda a bacia hidrográfica do rio Uruguai. A figura 15 – imagem “B” mostra a diferença entre os modelos LOGIT e LOGIT_m para *Leporinus*

obtusidens, sendo esta a única espécie que apresentou diferenças visíveis entre estes dois modelos. O modelo LOGIT_m estendeu a condição de presença da espécie até partes mais a montante que o modelo LOGIT nos locais de menor altitude da bacia hidrográfica. Em pequenos trechos do Alto Uruguai a situação se inverte, com o modelo LOGIT prevendo a ocorrência da espécie em locais mais a montante. Para as demais espécies (Figuras 9, 12 e 18 – letras “B”) as diferenças foram extremamente pequenas, onde as informações de presença ou ausência de cada um dos modelos diferem apenas em poucos metros na extensão dos rios. Tal informação, em conjunto com a análise dos gráficos de probabilidade de ocorrência (Figuras 4 – 7, imagens “B” e “D”), reforça o fato de que os modelos LOGIT e LOGIT_m são essencialmente iguais, e que proposição de Gonçalves (2010) não alterou significativamente a capacidade preditiva do modelo.

As diferenças entre os modelos LOGIT e LP foram mais acentuadas. De forma geral, o comportamento foi semelhante para todas as espécies: as maiores divergências encontram-se na parte baixa da bacia, onde o modelo LOGIT estende a predição de ocorrência até locais mais à montante nos pequenos segmentos de rio, resultando em distribuição prevista aparentemente superestimada, principalmente para o território argentino, não coberto pela matriz amostral. Ao mesmo tempo, identifica-se que o modelo LOGIT também estende as distribuições nos grandes rios de altitude, embora de forma menos marcante. Já nos segmentos intermediários da bacia os resultados foram similares, com o modelo LP apresentando maior de extensão de probabilidade de ocorrência para a maioria das espécies, exceto *Salminus brasiliensis*.

A Tabela 8 apresenta os valores de PMF_{Altitude} e PMF_{Área de Bacia} obtidos pelo modelo LOGIT_m no trabalho de Gonçalves (2010) e neste estudo, além dos dados do modelo LP para os mesmos parâmetros. Embora tenha sido aplicada a mesma base de dados nos dois trabalhos, o modelo LOGIT_m apresentou resultados bastante diferentes em cada um dos casos, onde Gonçalves (2010) obteve valores sempre mais elevados em relação aos obtidos neste trabalho. O modelo LP apresentou PMF_{Altitude} maiores que o modelo LOGIT_m, tanto no estudo de Gonçalves (2010) como nos testes feitos neste trabalho. Os PMF_{Área de Bacia} alcançados pelo modelo LP foram menores que os obtidos por Gonçalves (2010) com o modelo LOGIT_m. Tais dados ratificam a instabilidade apresentada pelo modelo LOGIT_m nas demais análises, mostrando as inúmeras soluções de mesma qualidade estatística a que o modelo pode chegar.

A Figura 19 mostra a distribuição de frequência dos pontos de presença conforme cada descritor ambiental. Os resultados de PMF alcançados com o modelo

LP estão contidos nos intervalos representados pelas barras de cor preta. A tendência apresentada pelo percentual de presença mostra que o modelo LP alcançou sempre resultados mais próximos do ponto médio de cada curva que o modelo LOGIT_m (Tabela 8; Figura 19).

O modelo proposto por Gonçalves (2010) teve por finalidade incluir na equação um parâmetro com significado biológico direto, como o Ponto de Mudança de Fase, ausente no modelo LOGIT. Entretanto, a possibilidade de que muitas combinações de Pontos de Mudança de Fase apresentem a mesma qualidade estatística de ajuste no modelo LOGIT_m faz com que não haja convergência ou estabilidade na estimativa deste parâmetro. Ou seja, o resultado alcançado depende das condições iniciais da análise. Neste sentido, não se considera precisa ou confiável a estimativa de Ponto de Mudança de Fase (PMF) no modelo LOGIT_m, não se justificando a aplicação do mesmo.

Os resultados obtidos ressaltam os estudos de Alves & Fontoura (2009) e Gonçalves (2010), mostrando que o uso em conjunto das variáveis geomorfológicas altitude e área de bacia é capaz de prever a distribuição de espécies migradoras, talvez em função do elevado conjunto de parâmetros colineares de efeito direto, como temperatura da água, vazão, velocidade de corrente ou turbidez (ALVES & FONTOURA, 2009). Além disso, a capacidade de se obter dados sobre essas duas variáveis ambientais remotamente torna viável o estudo em grandes escalas, como bacias hidrográficas inteiras, a exemplo deste trabalho.

A escolha sobre qual modelo utilizar é bastante complexa e não pode ser resumida ao resultado de poucos critérios estatísticos. Os métodos aplicados neste trabalho devem fazer parte de uma avaliação global do problema e a análise isolada de cada resultado deve ser evitada. A interpretação conjunta de todos os dados obtidos mostra que o modelo LP consegue melhores resultados que os demais modelos para prever a distribuição de peixes migradores, principalmente pela solução do problema de convergência e compensação encontrado nos modelos LOGIT e LOGIT_m. Embora tenha obtido resultados piores no critério AIC devido a sua complexidade e maior variância residual, as consequentes melhorias quanto a estabilidade estatística, formato do campo dimensional e melhor capacidade preditiva em situações-limite justificam sua utilização.

Os modelos apresentados, entretanto, não incorporam dados como a presença de barreiras naturais ou artificiais, fatores que influenciam diretamente a distribuição de peixes migradores. Tal fato, juntamente com a complexidade dos problemas biológicos, explicam a margem de erro obtida por todos os modelos. Gonçalves (2010)

identificou diversas barreiras durante o processo amostral. Entretanto, a base de dados GIS disponível não permite a incorporação de pequenos acidentes potencialmente importantes para a distribuição das espécies, pois a resolução do modelo numérico do terreno (imagem de altitude obtida por satélite em pixels de 92 x 92m) não é adequada para este tipo de análise (ALVES & FONTOURA, 2009; GONÇALVES, 2010).

Por outro lado, o modelo proposto também não incorporou a presença de barreiras artificiais, de forma que a distribuição prevista deve aproximar-se da distribuição histórica das espécies, antes da implementação de barramentos, subsidiando ações de planejamento e licenciamento ambiental e reintrodução das espécies nos trechos de rios onde estas já ocorriam no passado.

Tabela 8: Valores de PMF_{Altitude} e $PMF_{\text{Área de Bacia}}$ obtidos com o modelo $LOGIT_m$ por Gonçalves (2010) e neste trabalho e pelo modelo LP para todas as espécies em análise na bacia do rio Uruguai.

	Modelo	Pontos de Mudança de Fase	
		Altitude (m)	Área de Bacia (km ²)
<i>Salminus brasiliensis</i>	$LOGIT_m$ (Gonçalves, 2010)	445	1045
	$LOGIT_m$	231	320
	LP	672	166
<i>Prochilodus lineatus</i>	$LOGIT_m$ (Gonçalves, 2010)	359	1156
	$LOGIT_m$	151	101
	LP	516	101
<i>Leposinus obtusidens</i>	$LOGIT_m$ (Gonçalves, 2010)	523	1299
	$LOGIT_m$	231	177
	LP	651	83
<i>Pseudoplatystoma corruscans</i>	$LOGIT_m$ (Gonçalves, 2010)	299	3997
	$LOGIT_m$	68	709
	LP	509	1623

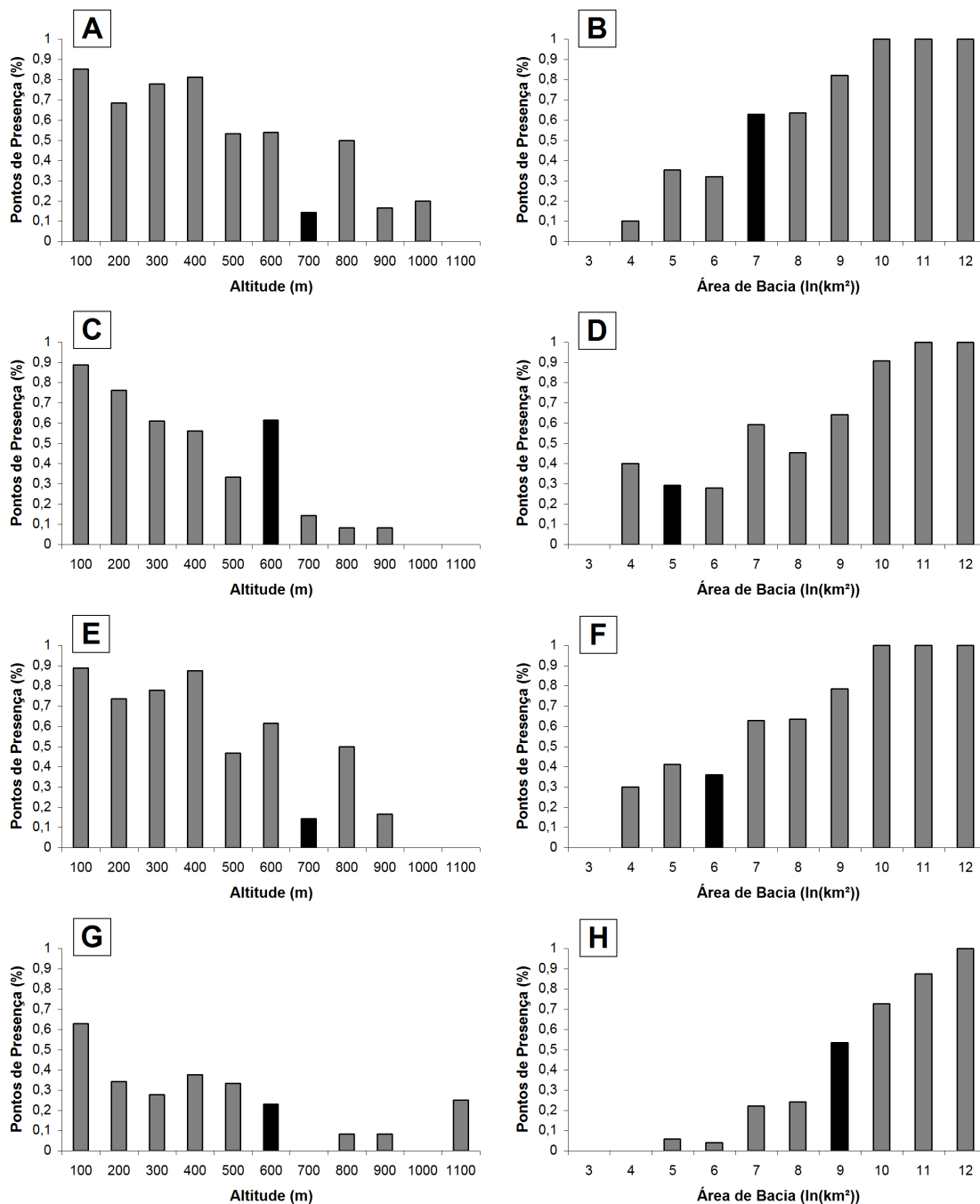


Figura 19: Distribuição de frequência dos pontos de presença conforme cada descritor ambiental para a bacia hidrográfica do rio Uruguai. “A” e “B” referem-se a *Salminus brasiliensis*, “C” e “D” referem-se a *Prochilodus lineatus*, “E” e “F” referem-se a *Leporinus otusidens* e “G” e “H” referem-se a *Pseudoplatystoma corruscans*. As barras em preto representam o intervalo que contém o Ponto de Mudança de Fase obtido com o modelo LP.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AGOSTINHO, C. S.; AGOSTINHO, A. A.; PELICICE, F.; ALMEIDA, D. A. DE; MARQUES, E. E. Selectivity of fish ladders: a bottleneck in Neotropical fish movement. **Neotropical Ichthyology**, v. 5, n. 2, p. 205-213, 2007.

ALVES, T. P. & FONTOURA, N. F. Statistical distribution models for migratory fish in Jacuí basin, South Brazil. **Neotropical Ichthyology**, v. 7, n. 4, p. 647-658, 2009.

ANDERSON, R. P., LEWC, D. e PETERSON, A. T. Evaluating predictive models of species' distributions: criteria for selecting optimal models. **Ecological Modelling**, v. 162 p. 211-232, 2003.

ANDRADE, P. M. & BRAGA, F. M. S. Reproductive seasonality of fishes from lotic stretch of the Grande River, High Paraná River Basin, Brazil. **Brazilian Journal of Biology**, v. 65, p. 387-394, 2005.

AUSTIN, M. P. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. **Ecological Modelling** v. 157, p. 101-118, 2002.

_____. Species distribution models and ecological theory: a critical assessment and some possible new approaches. **Ecological Modelling**, v. 200, p. 1-19, 2007.

BERTOLETTI, J. J. Contribuição ao conhecimento de passagens de peixes em barragens no Rio Grande do Sul. **Comunicações do Museu de Ciências e Tecnologia da PUCRS**, v. 42, p. 131-138, 1987.

BEWICK, V., CHEEK, L. e BALL, J. Statistics review 14: Logistic regression. **Critical Care**, v. 9, n. 1, p. 112-118, 2005.

BRAUNISCH, V., BOLLMANN, K., GRAF, R. F. e HIRZEL, A. H. Living on the edge - Modelling habitat suitability for species at the edge of their fundamental niche. **Ecological Modelling**, v. 214, p. 153-167, 2008.

BURNHAM, K. P. & ANDERSON, D. R. Multimodel Inference : Understanding AIC and BIC in Model Selection. **Sociological Methods & Research**, v. 33, p. 261-304, 2004.

Centrais Elétricas do Sul do Brasil S/A Florianópolis: Eletrosul, 1985. (Aqüicultura).

CAPELETI, A. R.; PETRERE JR, M. Migration of the curiMBatá *Prochilodus lineatus* at the waterfall " cachoeira de eMas " of the Mogi-guaçu riVer - são Paulo , Brazil *. **Brazilian Journal of Biology**, v. 66, p. 651-659, 2006.

CLARK, J. S., CARPENTER, S. R., BARBER, M., COLLINS, S., DOBSON, A., FOLEY, J. A., LODGE, D. M., PASCUAL, M., JR., R. P., PIZER, W., PRINGLE, C., REID, W. V., ROSE, K. A., SALA, O., SCHLESINGER, W. H., WALL, D. H. e WEAR, D. Ecological forecasts: an emerging imperative. **Science**, v. 293, p. 657-660, 2001.

CRAMER, J. S. **The origins and development of the logit model**. Cambridge, UK: Cambridge University Press, 2003. (Logit Models from Economics and Other Fields).

DTREG. Disponível em: <www.dtreg.com>. Acesso em: 21-maio-2010.

ELITH, J., PHILLIPS, S. J., HASTIE, T., DUDÍK, M., CHEE, Y. E. e YATES, C. J. A statistical explanation of MaxEnt for ecologists. **Diversity and Distributions**, v. 17, p. 43-57, 2011.

ESPÍNOLA, L. A., MINTE-VERA, C. V. e JÚLIO JR., H. F. Invasibility of reservoirs in the Paraná Basin, Brazil, to *Cichla kelberi* Kullander and Ferreira, 2006. **Biological Invasions**, 2009.

FRANKLIN, J. **Mapping Species Distributions - Spatial Inference e Prediction**. New York: Cambridge University Press, 2010. p. 320

GODINHO, A. L. & KYNARD, B. Migratory Fishes of Brazil: Life History and Passage Needs. **River Research and Applications**, v. 25, p. 702-712, 2009.

GODOY, M. P. A escada de peixes de Cachoeira de Ema, rio Mogi Guassú, estado de São Paulo, Brasil. **Comunicações do Museu de Ciências e Tecnologia da PUCRS**, v. 43, p. 139-151, 1987.

GONÇALVES, L. **Modelo estatístico de distribuição presumida de peixes migradores da bacia do rio Uruguai (RS/SC), sul do Brasil**. (2010). 48 f. Dissertação (Mestrado em Zoologia), Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2010.

GUISAN, A. & THUILLER, W. Predicting species distribution: offering more than simple habitat models. **Ecology Letters**, v. 8, p. 993-1009, 2005.

GUISAN, A. & ZIMMERMANN, N. E. Predictive habitat distribution models in ecology. **Ecological Modelling**, v. 135, p. 147-186, 2000.

HAIR JR, J., ANDERSON, R. E., TATHAM, R. L., BLACK, W., C. **Multivariate Data Analysis**. 4ª ed. New Jersey: Prentice-Hall, 1995.

HALL, C. A. S. Migration and Metabolism in temperate stream ecosystem. **Ecology**, v. 53, p. 585-604, 1972.

HOSMER, D. W. & LEMESHOW, S. **Applied Logistic Regression**. 2ª. ed. New York: John Wiley and Sons, 2000.

JOHNSON, J. B. & OMLAND, K. S. Model selection in ecology and evolution. **TRENDS in Ecology and Evolution**, v. 19, n. 2, 2004.

JUSTUS, J. O. Hidrografia. **Geografia do Brasil: Região Sul**. Rio de Janeiro: IBGE, 1990. Cap.190-218.

KATO, L. A. & BELLINI, M. Atribuição de significados biológicos às variáveis da equação logística: uma aplicação do cálculo nas ciências biológicas. **Ciência & Educação**, v. 15, n. 1, p. 175-188, 2009.

KELMA, J., PEREIRA, M. V. F., NETO, T. A. A. e SALES, P. R. H. Hidreletricidade. In: REBOUÇAS, A. C., BRAGA, B. & TUNDISI, J.G., (Ed.). **Águas Doces no Brasil: capital ecológico, uso e conservação**. São Paulo: Escrituras, 2002. p. 371-418.

KOHAVI, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. **Methodology**, 1995.

MASTRORILLO, S.; LEK, S.; DAUBA, F.; BELAUD, A. The use of artificial neural networks to predict the presence of small-bodied fish in a river. **Freshwater Biology**, p. 237-246, 1997.

LISCHKE, H., GUIBAN, A., FISCHLIN, A. e BUGMANN, H. Vegetation responses to climate change in the Alps-Modeling studies. In: CEBON, P. *et al* (Ed.). **A View from the Alps: Regional Perspectives on Climate Change**. Boston: MIT Press, 1998. p. 309-350.

LEBRETON, J. D., BURNHAM, K. P., CLOBERT, J. e ANDERSON, D. R. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. **Ecological Monographs**, v. 62, p. 67-118, 1992.

LOWE-MCCONNELL, R. H. **Estudos ecológicos de comunidades de peixes tropicais**. São Paulo: Edusp, 1999.

MARGULES, C. R. & AUSTIN, M. P. Biological models for monitoring species decline: the construction and use of data bases. **Philosophical Transactions of the Royal Society of London**, v. B 344, p. 69-75, 1994.

MASTRORILLO, S.; LEK, S.; DAUBA, F.; BELAUD, A. The use of artificial neural networks to predict the presence of small-bodied fish in a river. **Freshwater Biology**, p. 237-246, 1997.

MATTHEWS, W. J., HOUGH, D. J. e ROBISON, H. W. Similarities in fish distribution and water quality patterns in streams of Arkansas: congruence of multivariate analyses. **Copeia**, p. 296-305, 1992.

MEDEIROS, J. S. D. & CÂMARA, G. **Geoprocessamento para projetos ambientais**. São José dos Campos: Instituto Nacional de Pesquisas Ambientais, 2001.

MOILANEN, A., LEATHWICK, J. e ELITH, J. A method for spatial freshwater conservation prioritization. **Freshwater Biology**, v. 53, p. 577-592, 2008.

MOORE, C. H., HARVEY, E. S. e NIEL, K. P. V. Spatial prediction of demersal fish distributions: enhancing our understanding of species-environment relationships. **ICES Journal of Marine Science**, v. 66, p. 2068-2075, 2009.

MOURELL, C. & EZCURRA, E. Species richness of Argentine cacti: a test of biogeographic hypotheses. **Journal of Vegetable Science**, v. 7, p. 667-680, 1996.

PAIM, E. S. & ORTIZ, L. S. **Hidrelétricas na bacia do rio Uruguai: Guia para ONG's e movimentos sociais**. Porto Alegre: Núcleo Amigos da Terra Brasil, 2006. 80 p.

PEARCE, J. & FERRIER, S. Evaluating the predictive performance of habitat models developed using logistic regression. **Ecological Modelling**, p. 225-245, 2000a.

_____. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. **Ecological Modelling**, v. 128, p. 127-147, 2000b.

PEARSON, R. G. **Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis**. 2007. Disponível em: <<http://ncep.amnh.org>>. Acesso em: 8-set.-2010.

PHILLIPS, S. J., ANDERSON, R. P. e SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecological Modelling**, v. 190, p. 231-259, 2006.

PHILLIPS, S. J., DUDÍK, M., ELITH, J., GRAHAM, C. H., LEHMANN, A., LEATHWICK, J. e FERRIER, S. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. **Ecological Applications**, v. 19, p. 181-197, 2009.

PITCHER, T. J. & HART, P. J. B. **Fisheries Ecology**. Connecticut: The Avi Publishing Company Inc, 1982.

PORTER, M. S., ROSENFELD, J. e PARKINSON, E. A. Predictive models of fish species distribution in the Blackwater Drainage, British Columbia. **North American Journal of Fisheries Management**, v. 20, p. 349 - 359, 2000.

POUILLY, M., BARRERA, S. e ROSALES, C. Changes of taxonomic and trophic structure of fish assemblages along an environmental gradient in the Upper Beni watershed (Bolivia). **Journal of Fish Biology**, v. 68, p. 137-156, 2006.

STOCKWELL, D. & PETERS, D. The GARP modelling system: problems and solutions to automated spatial prediction. **International Journal of Geographical Information Science**, v. 13, n. 2, p. 143 - 158, 1999.

TURNER, W., SPECTOR, S., GARDINER, N., FLADELAND, M., STERLING, E. e STEININGER, M. Remote sensing for biodiversity science and conservation. **TRENDS in Ecology and Evolution**, v. 18, n. 6, p. 306-314, 2003.

VAZZOLER, A. E. A. M. **Biologia da reprodução de peixes teleósteos: Teoria e Prática**. São Paulo: Eduem, 1996.

WADDINGTON, C. H. **O homem e a ciência: instrumental para o pensamento**. São Paulo: Itatiaia, 1979.

WEBER, E., HASENACK, H. & FERREIRA, C. J. S. 2004. **Adaptação do modelo digital de elevação do SRTM para o sistema de referência oficial brasileiro e recorte por unidade da federação**. Porto Alegre, RS: Centro de Ecologia da UFRGS.

ZUCCHINI, W. An Introduction to Model Selection. **Journal of mathematical psychology**, v. 44, n. 1, p. 41-61, 2000.