

A MODELAGEM PREDITIVA NA ARQUEOLOGIA: UM OLHAR CRÍTICO A PARTIR DO ESTUDO DE CASO NO TERRITÓRIO XETÁ

PREDICTIVE MODELING IN ARCHEOLOGY: A CRITICAL LOOK FROM THE CASE STUDY IN XETÁ TERRITORY

Fabiana Terhaag Merencio

Como citar este artigo:

MERENCIO, Fabiana Terhaag. A modelagem preditiva na Arqueologia: um olhar crítico a partir do estudo de caso no território Xetá. Cadernos do Lepaarq, v. XVII, n.33, p. 25-45, Jan-Jun. 2020.

Recebido em: 17/01/2020

Aprovado em: 30/01/2020

Publicado em: 09/05/2020

ISSN 2316 8412

A modelagem preditiva na Arqueologia: um olhar crítico a partir do estudo de caso no território Xetá

Fabiana Terhaag Merencio ^a

Resumo: A modelagem preditiva arqueológica é um método importante para pesquisas preventivas e acadêmicas, pois tem como principal objetivo indicar áreas potenciais para presença de vestígios arqueológicos. Este artigo apresenta a aplicação da modelagem preditiva indutiva através da análise de regressão logística para identificar as variáveis ambientais relevantes para a presença dos acampamentos Xetá. Os resultados da modelagem preditiva são comparados com as informações etnográficas de organização do território Xetá, indicando-se problemas na aplicação desse método de forma acrítica.

Abstract: Archaeological predictive modelling is an important method for cultural resource management (CRM) and academic research, as its main objective is to indicate potential areas for the presence of archaeological remains. This paper presents the application of inductive predictive modeling through logistic regression analysis to identify the relevant environmental variables for the presence of Xetá camps. The results of the predictive modeling are compared with the ethnographic information of Xetá territory organization, indicating problems in the application of this method in an uncritical way.

Palavras Chave:

Análise espacial, Modelagem preditiva, Regressão Logística, Xetá.

Keywords:

Spatial analysis, Predictive modeling, Logistic regression, Xetá.

^a Doutoranda do Programa de Pós-graduação em Arqueologia do Museu de Arqueologia e Etnologia da Universidade de São Paulo (PPGARq/MAE/USP), Brasil. Apoio financeiro Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Brasil. E-mail: f.terhaag@gmail.com

INTRODUÇÃO

As análises espaciais foram intensificadas na Arqueologia ao longo da década de 1970, sob a influência da Escola de Geografia Antropológica Alemã e da Nova Geografia, destacando-se a Arqueologia Espacial na Inglaterra, com os trabalhos de Clarke (1972), Watson, LeBlanc e Redman (1971) e Hodder & Orton (1976), e a Nova Arqueologia nos Estados Unidos, com a pesquisa de Willey (1953) (ROOD, 1982; GAMBLE, 1987). Métodos e técnicas desenvolvidas nesse período e baseadas em modelos estatísticos e quantitativos são empregadas até os dias atuais na disciplina (GILLINGS & WHEATLEY, 2005; HU, 2012).

A modelagem preditiva, foco deste artigo, é uma das análises espaciais mais aplicadas na disciplina através de Sistemas de Informação Geográfica, sobretudo em pesquisas preventivas. O objetivo da análise é predizer áreas com potencial para presença de vestígios arqueológicos, tendo como base: a correlação entre registro arqueológico e variáveis ambientais e culturais (técnica indutiva) ou, a formulação de hipóteses sobre o comportamento humano que influenciaram na escolha de uma determinada área (técnica dedutiva) (KOHLER & PARKER, 1986; GILLINGS & WHEATLEY, 2005; VERHAGEN, 2007).

A aplicação do método é considerada importante tanto pelo poder público, pois possibilita a identificação de áreas de interesse de proteção do patrimônio arqueológico, como por pesquisas preventivas, pois permite reduzir os custos e tempo na prospecção de grandes áreas (KOHLER & PARKER, 1986; VERHAGEN, 2007). No Brasil, com a publicação da Instrução Normativa nº1/2015 do Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN), passou a ser solicitado a inclusão de “modelos preditivos de avaliação, de forma a justificar claramente os locais onde deverão ser realizadas as prospecções” (BRASIL, 2015, p. 12).

Este artigo apresenta os procedimentos para aplicação da modelagem preditiva indutiva através da regressão logística na área e território Xetá, no noroeste do Paraná, contrapondo os resultados obtidos com as informações de mudança da organização territorial durante a intensificação dos conflitos interétnicos. O objetivo é apontar problemas no uso acrítico da modelagem preditiva para identificar locais com alto potencial para presença de vestígios arqueológicos, indicando-se possíveis soluções e cuidados para aplicação do método.

Área de pesquisa

Os Xetá são um grupo indígena que ocupavam a região noroeste do Paraná, na Serra dos Dourados, margem esquerda do rio Ivaí. Em termos linguísticos, o Xetá foi associado ao ramo meridional, sub-ramo I, da família Tupi-Guarani (RODRIGUES, 1978, 1985). Os primeiros relatos de contato são de expedições que percorreram a região do Rio Ivaí no século XIX (KELLER & KELLER,

1866; BIGG-WHITER, 1878; MOTA, 2013). As notícias dos contatos foram intensificadas entre 1940-60, com o avanço das frentes colonizadoras na região, que passaram a registrar acampamentos abandonados (FERNANDES, 1959b; SILVA, 1998).

As primeiras expedições enviadas pelo Serviço de Proteção aos Índios (SPI) para verificar tais informações ocorreram em 1945 e 1949 (KOZÁK et al., 1981). A presença de um grupo Xetá na Fazenda Santa Rosa em 1954, motivou o envio de mais três expedições em 1955 (Figura 1A). Nos anos seguintes, entre 1956 a 1961, José Loureiro Fernandes da Universidade Federal do Paraná, organizou o envio de expedições para registrar informações (LAMING-EMPERAIRE et al., 1978). Entre 1945 a 1961, foram registrados pelas expedições 21 acampamentos em uma área de 967 km². Segundo os sobreviventes Xetá, o território ocupado por seu grupo era muito maior, abrangendo o rio Ivaí e seus afluentes, até o rio Piquiri, em uma área estimada de 28mil km² (SILVA, 1998). Pesquisas da época descreveram os Xetá como um grupo de caçadores-coletores de alta mobilidade, com permanência máxima de três dias nos acampamentos, distribuídos em pequenas famílias nucleares e população estimada de 100-300 pessoas (FERNANDES, 1959a; LAMING-EMPERAIRE et al., 1978).

Ressalta-se que a alta mobilidade identificada era de fato um produto da adaptação do grupo frente as pressões externas. Segundo os sobreviventes Tuca, Kuein e Tikuein, até o início de 1950 havia três tipos de ocupações (SILVA, 2003): (1) aldeia grande (*Óka awatxu*) ocupada pelo grupo local, composto por várias famílias nucleares, por seis meses no verão, período em que celebravam rituais; (2) acampamento de habitação (*Óka kã*) ocupados por famílias nucleares durante o inverno, na temporada de caça, por dois a três meses; e (3) os barreiros (*'ogwa kã*), que eram acampamentos de caça e preparação de carne localizados em pontos estratégicos próximos às margens de rios e córregos (Figura 1B). No período em que o contato foi intensificado, apenas os acampamentos de habitação eram utilizados pelo grupo, concentrando todas as atividades, exceto os rituais que não eram mais praticados. Os sobreviventes indicam que a mudança do modelo de assentamento e organização do território ocorreu devido aos conflitos interétnicos provocados por pressões externas de grupos indígenas (Guarani e Kaingang) e avanço das frentes de colonização.

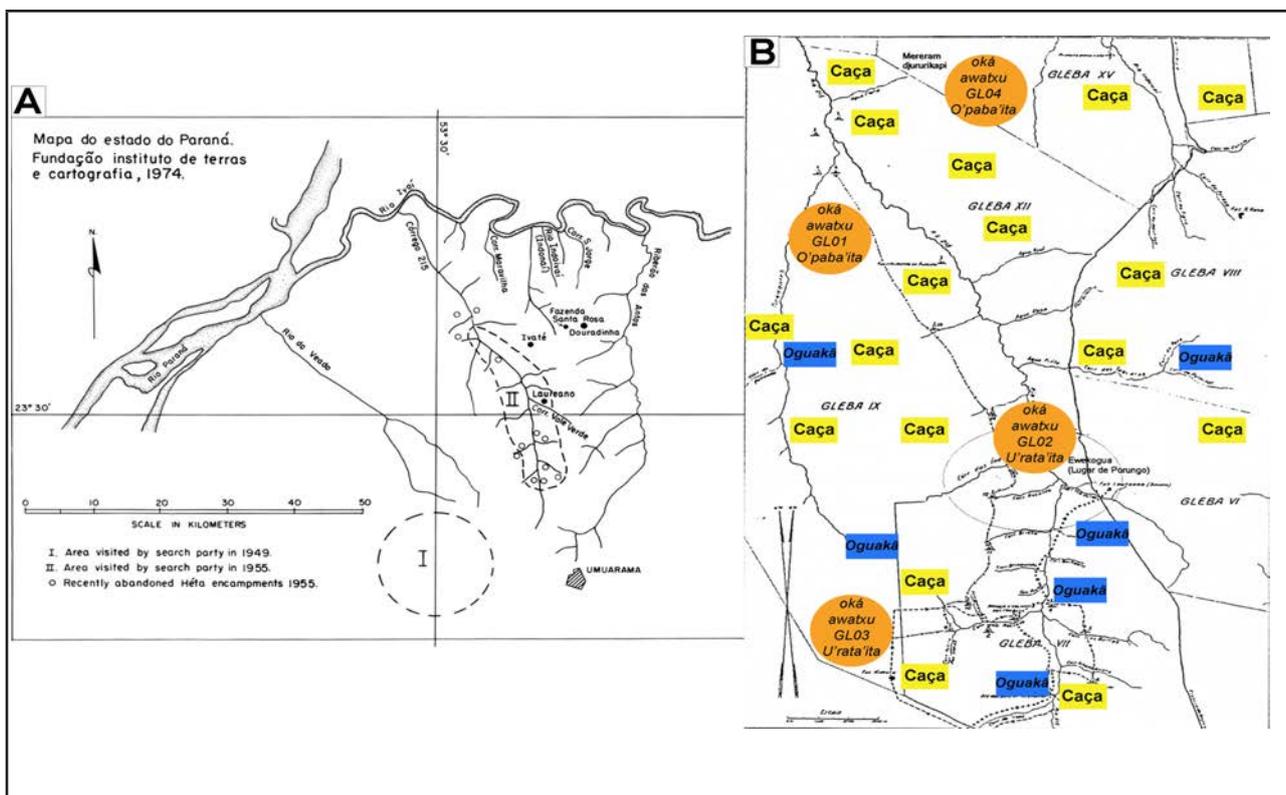


Figura 1: (A) Mapa com a indicação das áreas percorridas pelas expedições de 1949 e 1955 (KOZÁK et al., 1979, p. 359).
 (B) Localização das aldeias grandes (laranja), áreas de caça e barreiros (SILVA, 2003, p. 173, adaptado).

Sistemas de Informação Geográfica na Arqueologia

Os Sistemas de Informação Geográfica (SIG) foram desenvolvidos de formas independentes a partir de 1960 por diferentes organizações e suas origens são difíceis de traçar (COPPOCK & RHIND, 1991). Um SIG é composto por diferentes componentes, a saber: 1) dados geográficos, obtidos de forma primária (levantamentos de campo) ou secundária (arquivos raster/vetoriais¹); 2) métodos de análise dos dados, como procedimentos específicos para responder uma questão; 3) softwares para coleta, armazenamento, processamento e análise dos dados; e 4) hardware para funcionamento dos softwares (ROSA, 2005).

Em um SIG, as informações geográficas são inseridas, armazenadas e visualizadas de forma distinta dos mapas tradicionais: enquanto um mapa apresenta de forma conjunta dados geográficos, como elevação, hidrografia, vegetação, limites geográficos e sítios arqueológicos, no SIG essas mesmas informações são desmembradas em camadas individuais e temáticas (layers) representadas por arquivos vetoriais ou raster. O uso de camadas temáticas facilita a organização e gerenciamento dos dados geográficos, permitindo que diferentes informações sejam integradas

¹ Raster apresentam dados espaciais contínuos em forma de células, como elevação do terreno, enquanto vetores representam dados espaciais não contínuos através de pontos, linhas e polígonos, como limites territoriais (WHEATLEY & GILLINGS, 2002).

no banco de dados, desde que mantendo as relações espaciais em relação à projeção utilizada (GILLINGS & WHEATLEY, 2005).

Embora o uso de técnicas cartográficas e análises espaciais na Arqueologia tenha sido intensificado na década de 1970, os primeiros trabalhos com SIG foram realizados a partir de 1980, em projetos de pesquisas regionais e desenvolvimento de modelos preditivos, verificando-se na década seguinte um impulso em análises espaciais na disciplina através do SIG, com a melhoria de interface de softwares e menor custo de hardware (GILLINGS & WHEATLEY, 2005; HU, 2012). O SIG é considerado uma poderosa ferramenta na Arqueologia pois possibilita a análise de uma grande quantidade de informações espaciais com múltiplos enfoques, como: construção de bases de dados regionais, gerenciamento do patrimônio arqueológico, e em especial, a realização de análises espaciais (WHEATLEY & GILLINGS, 2002; CONOLLY & LAKE, 2006).

Modelagem Preditiva na Arqueologia

As raízes da modelagem preditiva arqueológica (MPA) é anterior à popularização do SIG, sendo associada aos estudos de padrão e sistemas de assentamentos de 1950 (WILLEY, 1953) que exploravam a ideia de que o ambiente era um fator importante para a localização dos sítios arqueológicos, e também a partir de 1960 com o uso de análises quantitativas para investigar tais relações. O desenvolvimento da MPA foi intensificado nos Estados Unidos com a expansão das pesquisas preventivas e aprovação da Lei Nacional de Preservação Histórica de 1966, que promoveu a identificação, preservação e registro de sítios arqueológicos, impulsionando a realização de pesquisas preditivas, que passaram a ser realizadas com maior intensidade a partir de 1970 (VERHAGEN, 2007).

A utilização dos resultados para otimizar a etapa de prospecção tem sido considerada problemática, pois usualmente são priorizadas áreas que apresentam alto potencial, resultando na coleta de dados semelhantes aos da amostra inicial, sendo portanto ineficiente quando o objetivo é compreender padrões de sistemas de assentamento (VERHAGEN & WHITLEY, 2012). Nessa discussão, Wheatley (2004, p. 6) aponta três problemas no uso da MPA: 1) a correlação preditiva é considerada reducionista pois enfatiza características ambientais e exclui o comportamento humano como fator explicativo; 2) o comportamento humano não é resultado unicamente de estímulos ambientais; e 3) o uso de dados ambientais atuais, desconsidera que as paisagens arqueológicas são resultado de diferentes processos históricos.

É válido ressaltar que modelos são hipóteses formuladas que buscam explicar um universo empírico observado, ao mesmo tempo que possui caráter preditivo (KIPNIS, 1997). MPA, em especial, refere-se a formulação de hipóteses sobre a localização de sítios partindo do pressuposto de que um tipo particular de sítio tem a propensão a ocorrer em ambientes similares, tendo com objetivo de predizer áreas com potencial para identificação de vestígios desconhecidos (KOHLENER & PARKER, 1986; GILLINGS & WHEATLEY, 2005). Ressalta-se que outros métodos podem

ser realizados para se atingir o mesmo objetivo, como por exemplo, em análises de mobilidade para identificar possíveis áreas de ocupação². Apesar de ser possível tal uso, ressalta-se que as premissas e procedimentos de cada tipo de análise são distintos.

Para elaboração da MPA são utilizadas duas técnicas consideradas dicotômicas: a dedutiva (theory-driven) e a indutiva (data-driven). A dedutiva tem como base a formulação de hipóteses sobre a localização dos sítios, como por exemplo, proximidade de água ou elevação. Para cada variável considerada relevante, atribui-se diferentes pesos através de uma equação matemática. A indutiva parte de dados empíricos para verificação das correlações entre sítios arqueológicos e variáveis ambientais e culturais, através da aplicação de testes estatísticos, como a regressão logística, podendo-se extrapolar os resultados para uma área maior (KOHLENER & PARKER, 1986; WHEATLEY & GILLINGS, 2002; VERHAGEN & WHITLEY, 2012).

A principal diferença entre ambas está no método em si, pois os dados utilizados na técnica indutiva foram obtidos a partir de orientações teóricas, enquanto a formulação de hipóteses na técnica dedutiva, está intrinsecamente relacionada com a observação de dados empíricos (VERHAGEN, 2007; VERHAGEN & WHITLEY, 2012). Para construção do MPA na bacia do Rio Capivaras na encosta sul de Santa Catarina, Kozłowski & DeBlasis (2019) aplicaram as duas técnicas, e apesar das semelhanças visuais, indicaram que o modelo indutivo apresentou estatisticamente maior precisão e acurácia para o estudo de caso apresentado. No Brasil, outros exemplos de MPA são encontrados em: Kipnis (1997), Tsuchiya et al. (2002), Fonseca (2013; 2018), Kozłowski (2018), Perez (2018) e Perez et al. (2018).

Métodos e materiais

A construção da MPA consiste em quatro etapas: coleta de dados, análise estatística, aplicação do modelo e validação do mesmo (CONOLLY & LAKE, 2006, p. 181). Neste item são descritos os procedimentos para as três primeiras etapas, e a validação do modelo é discutida nos resultados. Todos os procedimentos de produção de mapas foram realizados no ArcGIS 10.2.2 e cálculos de regressão logística no Jamovi 1.1.5 (JAMOMI, 2019). Primeiramente, foram definidas as variáveis dependentes e independentes. A variável dependente, os acampamentos, é dicotômica, sendo atribuído valores de 1 para a presença e 0 para presença desconhecida. As coordenadas geográficas dos acampamentos Xetá foram obtidas através do georreferenciamento do mapa de Ney Barreto (FERNANDES, 1962); para os pontos aleatórios, foi gerado um vetor com 21 pontos; e os limites da área ocupada entre 1940-60 e respectivo território, foram estimados a partir dos mapas de Silva (1998) (Figura 2). As variáveis independentes selecionadas foram: elevação, vegetação, solo, declividade, custo da distância da água e proximidade de água. A Tabela 1 sintetiza os dados utilizados e procedimentos executados no ArcGIS através da Toolbox.

² Exemplos de análises de rotas de mobilidade: D'Ávila (2017), Silva (2018), Kozłowski & Corteletti (2019), e Milheira et al. (2019).

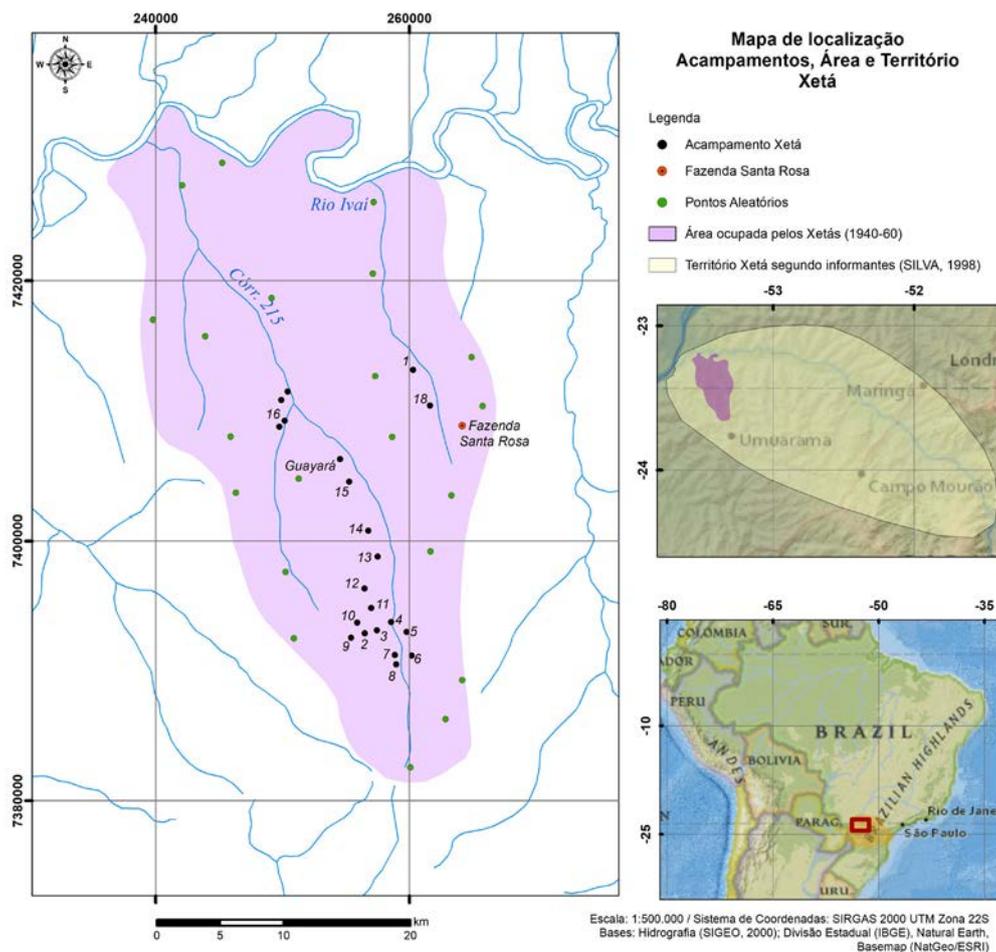


Figura 2: Mapa de localização dos acampamentos e pontos aleatórios na área ocupada pelos Xetá entre 1940-1960 e da área estimada para do território do grupo. A numeração dos acampamentos segue as informações do mapa de Ney Barreto.

Tabela 1: Dados geográficos

Variável	Classe	Tipo	Descrição	Comandos ArcGIS (via Toolbox)
Acampamentos e Pontos aleatórios	Shape	Primário	Pontos binários identificando presença (1) e presença não confirmada (0) de vestígios Xetá	<i>data management > feature class > create random points > constraining feature class (limite da área de pesquisa) > long (número de pontos a ser criado)</i>
Elevação (m)	Raster	Primário	Modelo digital de elevação (DEM), resolução de 30m. NASA/SRTMGL1.v3(2013)	
Declividade (°)	Raster	Secundário	Declividade da superfície em graus, gerado a partir do DEM.	<i>spatial analyst > surface > slope > input raster (elevação) > output measurement (graus) > z factor (1)</i>
Hidrografia	Shape	Primário	Massa de água e drenagens, escala 1:250.000. (IPARDES/ITCG)	

Variável	Classe	Tipo	Descrição	Comandos ArcGIS (via Toolbox)
Custo de distância da água	Raster	Secundário	Custo energético de distância da água em metros	<i>spatial analyst tools > distance > cost distance > input raster or feature source data (shape hidrografia) > input cost raster (declividade) > maximum distance (não preencher)</i>
Proximidade de água	Raster	Secundário	Distância euclidiana linear da água.	<i>spatial analyst > distance > euclidean distance > input raster or feature source data (shape hidrografia) > maximum distance (não foi indicado distância máxima) > output cell size (30= resolução raster elevação)</i>
Vegetação	Shape	Primário	Fitogeográfico do Paraná (1989/1990). (MAACK/ITCG/IBGE)	
Vegetação	Raster	Secundário	Dados contínuos criado a partir do vetor da vegetação	<i>Conversion tools > to raster > polyline to raster > value field (tipo de vegetação) > cell size (30= resolução raster elevação)</i>
Solo	Shape	Primário	Solos do estado do Paraná. (EMBRAPA/EMATER)	
Solo	Raster	Secundário	Dados contínuos criado a partir do vetor do solo.	<i>Conversion tools > to raster > polyline to raster > value field (tipo de solo) > cell size (30= resolução raster elevação)</i>

Os valores numéricos das variáveis independentes foram extraídos para os vetores de acampamentos e pontos aleatórios através da ferramenta *extract multi values to points*, localizada na Toolbox do ArcGIS (*spatial analyst tool > extraction*). Para realizar os cálculos da regressão logística, os dados geográficos obtidos foram exportados para tabela (*conversion tools > excel > table to excel*) (Material suplementar 1).

A partir desses dados, a MPA pode ser realizada tanto pela técnica dedutiva como indutiva. A aplicação da primeira usualmente não indica a influência que cada variável possui, nem um parâmetro de avaliação do modelo em relação à distribuição dos sítios observados (GILLINGS & WHEATLEY, 2005, p. 396). Tais questões são melhor respondidas através da aplicação de análises de regressão linear e logística, que possibilitam estimar a relação que cada variável independente possui para a presença da variável dependente. A regressão linear apresenta resultados em valores absolutos, enquanto a regressão logística proporciona resultados em termos de probabilidades, sendo mais interessante para a modelagem preditiva (WHEATLEY & GILLINGS, 2002; RIRIS, 2010).

“O modelo de funcionamento da regressão logística é dado pela equação: $L=B_0+(B_1X_1)+(B_2X_2)+(B_3X_3)...(B_nX_n)$, onde B_0 corresponde ao coeficiente da variável dependente; $B_1...n$ ao coeficiente das variáveis independentes, cujo valor indica o peso que cada variável independente possui; e X_n aos valores das variáveis relevantes (RIRIS, 2010). Para estimar a probabilidade, a equação acima é adaptada da seguinte forma (KVAMME, 1988):

$$P = \frac{1}{(1 + \text{Exp}(-(\beta_0 + (\beta_1X_1) + (\beta_2X_2) + (\beta_3X_3) + \dots (\beta_nX_n))))}$$

Para análise da regressão logística, os dados geográficos obtidos no ArcGIS foram inseridos no Jamovi 1.1.5 (JAMOVI, 2019) (*regression > logistic regression > 2 outcomes/binomial*). A Tabela 2 apresenta os resultados do teste de razão de verossimilhança a partir do chi-quadrado (χ^2) aplicado para estimar o quanto cada variável independente é capaz de prever a variável dependente (HOLTON, 2014). Os resultados foram avaliados a partir do nível de confiança de 95%, onde os valores indicados na coluna p devem ser menores de 0.05. Em termo gerais, valores de p são utilizados para rejeitar ou aceitar a hipótese nula, onde não há uma correlação entre as variáveis (ALDENDERFER, 2005).

Ao total foram realizadas seis etapas de cálculos, excluindo-se em cada uma as variáveis com maior valor p, na seguinte sequência: declividade (p=0,837), solo (p=0,702), elevação (p=0,399), custo distância da água (p=0,288) e vegetação (p=0,081). A regressão logística indicou que a distância da água possui alta correlação para a presença dos acampamentos Xetá, com um valor de p=0,001, observando-se que os pontos estão localizados a uma distância média de 344 metros dos cursos d'água, sendo a máxima de 1,2km.

Tabela 2: Eliminação progressiva das variáveis independentes

	Variável	χ^2	df	p
Etapa 1	Elevação	0,5535	1	0,457
	Declividade	0,0422	1	0,837
	Custo distância água	1,7491	1	0,186
	Distância água	3,8978	1	0,048
	Solo	0,1417	1	0,707
	Vegetação	1,0528	1	0,305
Etapa 2	Elevação	0,608	1	0,436
	Custo distância água	1,799	1	0,180
	Distância água	4,177	1	0,041
	Solo	0,147	1	0,702
	Vegetação	1,076	1	0,300
Etapa 3	Elevação	0,711	1	0,399
	Custo distância água	1,685	1	0,194
	Distância água	4,045	1	0,044
	Vegetação	1,097	1	0,295
Etapa 4	Custo distância água	1,13	1	0,288
	Distância água	3,35	1	0,067
	Vegetação	2,98	1	0,084
Etapa 5	Distância água	6,98	1	0,008
	Vegetação	3,04	1	0,081
Etapa 6	Distância água	12,3	1	0,001

Finalizada a análise estatística, a equação de probabilidade da regressão logística foi inserida no raster calculator no ArcGIS (*Toolbox>spatial analyst>map algebra*), com a seguinte sintaxe: $1 / (1 + \text{Exp}(- (B_0 + (B_1 * X_1))))$. Os valores dos coeficientes são informados na tabela

Model Coefficients (coluna Estimate) gerada na regressão logística, onde o Intercept corresponde à variável dependente (Tabela 3). Para exemplificar a equação de probabilidade com os valores dos coeficientes obtidos, foi inserida a seguinte fórmula: $1 / (1 + \text{Exp} (- (1.45194 + (-0.00258 * \text{"raster distância água"}))))$.

Tabela 3: Coeficientes do modelo

Predictor	Estimate	SE	Z	p
Intercept	1.45194	0.578	2.51	0.012
Distância água	-0.00258	0.000908	-2.84	0.004

RESULTADOS E DISCUSSÃO

Os resultados obtidos no MPA para a área ocupada pelos Xetá entres 1940-60 foram extrapolados para o território ocupado pelo grupo. Na Figura 3A verifica-se a presença de sítios líticos nas porções de alta probabilidade na área ocupada pelos Xetá. Os relatos das expedições que percorreram a região em 1842 e 1845 indicam a presença de grupos indígenas às margens do Rio Ivaí. Mota (2013b, p. 17-20) aponta que a descrição do uso de batoques de resina sugere que se tratava de grupos Xetá, contudo, o autor observa que esses vestígios também podem ser associados aos Kaiowá que ocupavam a margem esquerda do Rio Ivaí.

Em relação ao território estimado (Figura 3B), também observa-se a presença de sítios líticos em locais de alta probabilidade, com destaque para a porção sudeste. Nessa área há relatos da presença de grupos Xetá entre 1850 e 1870 nas proximidades da Colônia Tereza Cristina (MOTA, 2013, p. 45–47), e em 1872 na região do Salto do Ariranha, conforme Thomas Bigg-Whiter (1878). A partir das descrições de Bigg-Whiter, Kozáck et al. (1981) e Silva (1998, p. 234) apontam que o grupo descrito possivelmente tratava-se de antepassados dos Xetá.

A presença de sítios líticos na área ocupada pelos Xetá e no respectivo território, sugerem que tais locais possam ter alguma relação com o grupo. Contudo, é preciso cautela com tais conclusões; para se confirmar tais hipóteses, somente uma análise comparativa dos vestígios registrados nos sítios e da coleção de artefatos Xetá coletados pelas expedições etnográficas. Além disso, enfatiza-se que a extrapolação do modelo preditivo para todo o território pode ser perigosa, pois parte do pressuposto de que porções distintas da paisagem foram ocupadas de modos semelhantes.

Visualmente, o modelo gerado para a área ocupada apresenta alta precisão, com a distribuição da maior parte dos pontos na área de alta probabilidade (Figura 3A). A tabela de classificação obtida na regressão logística indica a quantidade de pontos preditos em cada categoria: do total de 21 acampamentos, 17 foram preditos corretamente (acurácia de 81%), enquanto 14 pontos aleatórios foram preditos (acurácia de 66,7%). Em suma, foram preditos corretamente 31 pontos do total de 42 analisados, resultando em uma acurácia geral do modelo de 73,8%.

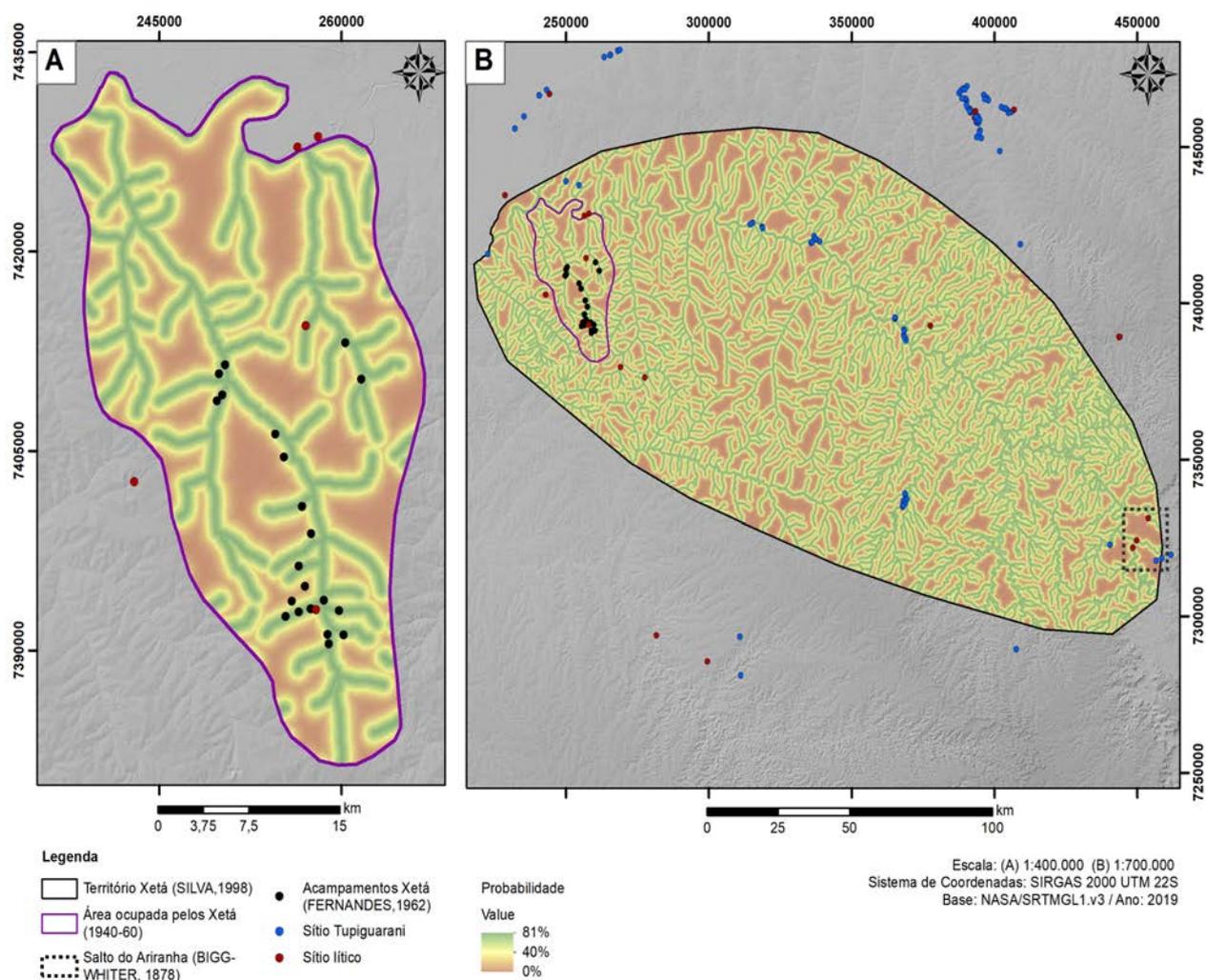


Figura 3: Mapa de probabilidade para presença de acampamentos Xetá: (A) território estimado; (B) área ocupada pelo grupo.

O cálculo de ganho estatístico de Kvamme (1988) é outra forma de se avaliar o modelo, através da fórmula: $G = 1 - \frac{Pa}{Ps}$, onde Pa é a porcentagem de área de um nível de probabilidade no modelo e Ps é a porcentagem de pontos registrados em cada nível. Para se obter os dados necessários para esse cálculo (Tabela 4), foram realizados os seguintes procedimentos no ArcGIS: 1) reclassificação do raster de probabilidade da área ocupada entre 1940-1960 em três níveis de probabilidade com intervalo igual (*spatial analyst>reclass>reclassify*); 2) cálculo da área de cada nível de probabilidade (*spatial analyst>zonal>zonal geometry as table*); 3) identificação do nível de probabilidade associado a cada ponto (*spatial analyst tool > extraction > extract multi values to points*). Pa/Ps

No cálculo do ganho estatístico, os resultados são apresentados em uma escala de 0 a 1, onde 0 indica que o modelo possui baixo desempenho e 1 alto desempenho, podendo também

apresentar valores negativos para áreas de baixa probabilidade (KVAMME, 1988; KOZLOWSKI, 2018). Os resultados indicam ganho mediano (0,55) para o nível de alta probabilidade, com baixa precisão (34%) e alta acurácia (76%). A presença de acampamentos nos níveis de baixa e média probabilidade, reflete uma baixa acurácia e precisão do modelo, apesar do ganho estatístico negativo (Tabela 4) (KVAMME, 1988; KOZLOWSKI, 2018).

Tabela 4: Cálculo ganho estatístico

Probabilidade	Área (km ²)	Acampamentos	Precisão (% área)	Acurácia (% acampamentos)	Ganho (=1-(Precisão/Acurácia))
Baixa (0-27%)	377,02	2	39%	10%	-3,093298
Média (27-54%)	264,49	3	27%	14%	-0,914375
Alta (54-81%)	325,61	16	34%	76%	0,5581074
Total	967,12	21	100%	100%	

Em termos estatísticos, considera-se que o modelo preditivo gerado neste estudo de caso apresentou resultados satisfatórios, contudo, é simples na medida em que apresenta falhas assim como outros modelos preditivos estimados somente com base em variáveis ambientais e que se utilizam de banco de dados de sítios arqueológicos inadequados. Por exemplo, hipoteticamente o modelo apresentado indica o potencial para presença de acampamentos de habitação (*Óka kã*) e possivelmente de alguns barreiros (*’ogwa kã*), e não contemplaria outras categorias de ocupações como as aldeias grandes (*Óka awatxu*). A avaliação crítica do modelo aponta que: 1) alguns dos acampamentos apresentam características de implantação distintas do restante, e desse modo, o modelo não indica a probabilidade com alta precisão, apesar da alta acurácia, ou 2) alguns dos pontos considerados como acampamentos, possuíam uso distinto, e portanto, apresentam características distintas.

A MPA indutiva apresentada partiu propositalmente da avaliação de variáveis ambientais, usualmente utilizadas em pesquisas preditivas devido a facilidade de acesso aos dados. A inclusão de variáveis culturais, como visibilidade, podem resultar em modelos preditivos mais robustos, e recomenda-se que essas sejam incluídas (VERHAGEN et al., 2013).

Em pesquisa preventivas, MPA é um método importante para identificação de sítios arqueológicos, prevenindo a destruição desses em obras de grande impacto. A grande questão sobre MPA reside em como os resultados são utilizados: se os considerarmos como modelos explicativos de padrões e sistemas de assentamentos, corre-se o risco de formular generalizações simplistas sobre o comportamento humano e suas relações com a paisagem; se os utilizarmos para testar a localização de sítios arqueológicos, então os modelos preditivos serão empregados como ponto de partida para formulação das estratégias de pesquisa, o que parece inadequado nos casos em que dados de uma área são extrapolados para uma região afastada e com poucos dados arqueológicos disponíveis (VERHAGEN & WHITLEY, 2012). Esses problemas são causados, em geral, pelo uso inadequado da

escala de pesquisa e dos dados utilizados, como uso generalizado dos sítios de uma região ou banco de dados insuficientes.

Nesse sentido, enfatiza-se que modelos preditivos são formulações hipotéticas sobre a localização de sítios arqueológicos, e não devem ser considerados como um instrumento que solucionará os problemas de prospecções de grandes áreas. É preciso utilizar a modelagem preditiva de forma crítica, tendo claro as limitações e falhas que o método possui e que se trata de uma “ferramenta auxiliar ao longo do processo” de prospecção de grandes áreas (KOZLOWSKI & DEBLASIS, 2019, p. 89). Para tal, é importante estabelecer o tipo de fenômeno que se pretende prever, gerando modelos específicos para cada categoria. Isso é possível a partir de uma simples classificação da amostra de sítios em diferentes categorias (seja funcional, temporal ou características da cultura material), pois caso contrário, o resultado será um modelo altamente generalizante e pouco explicativo. Nos casos em que a modelagem preditiva será utilizada como instrumento para guiar a seleção de áreas de prospecção, além da geração de modelos específicos, recomenda-se a verificação amostral das áreas de baixo-médio potencial, a partir de uma reavaliação em campo das características básicas da área de pesquisa, como acessibilidade, visibilidade ou presença de fontes de matérias-primas.

CONSIDERAÇÕES FINAIS

Neste artigo foram apresentados os procedimentos detalhados para realização da MPA indutiva através da análise de regressão logística na área ocupada pelos Xetá entre 1940-60, extrapolando-se os resultados para o território do grupo. A aplicação do método apontou como variável relevante para presença dos acampamentos a distância da água. A validação do modelo em termos estatísticos, indicou que foram preditos corretamente 81% dos acampamentos na área Xetá, enquanto o cálculo de ganho estatístico apresentou valor mediano (0,55) para as áreas de alta probabilidade, indicando a presença correta de 76% dos acampamentos.

Os resultados da MPA foram comparados com as informações etnográficas de organização do território Xetá, indicando-se problemas na aplicação do método de forma acrítica. Foram enfatizados dois pontos principais: 1) a extrapolação do modelo preditivo obtido na área ocupada entre 1940-60 para todo o território pode ser perigosa, pois considera que porções distintas da paisagem foram ocupadas de modos semelhantes; e 2) o modelo gerado não prediz a presença de categorias de ocupações não contempladas a priori na amostra, como as aldeias grandes (*Óka awatxu*).

A partir disso, foi apresentada uma breve discussão sobre a forma como os resultados da modelagem preditiva são utilizados e respectivos problemas, provocados em grande parte pelo uso inadequado da escala em que o modelo é extrapolado, e da qualidade dos dados utilizados. Considera-se que a MPA é um método importante para o registro de sítios arqueológicos em

MERENCIO, Fabiana Terhaag. A modelagem preditiva na Arqueologia: um olhar crítico a partir do estudo de caso no território Xetá. Cadernos do Lepaarq, v. XVII, n.33, p. 25-45, Jan-Jun. 2020.

pesquisas preventivas, desde que alguns cuidados sejam considerados com a escala da aplicação do modelo e fenômenos analisados. Se aplicado tendo tais preocupações em mente, a MPA pode apresentar resultados para formulação de generalizações testáveis sobre a distribuição espacial de sítios arqueológicos.

Agradecimentos

Ao Henrique de Sena Kozlowski, Glauco Pérez e Thiago Umberto Pereira pelas discussões sobre modelos preditivos. Este artigo é uma extensão da pesquisa de mestrado realizada no PPGAA/UFPR, e contou com apoio financeiro da CAPES.

REFERÊNCIAS

- ALDENDERFER, Mark. Statistics for Archaeology. IN: MASCHNER, Herbert D.G.; CHIPPINDALE, Christopher. Handbook of Archaeological Methods. Walnut Creek, CA: Altamira Press, 2005. p. 501–553.
- BIGG-WHITER, Thomas. Novo caminho no Brasil meridional: a província do Paraná, três anos de vida em suas florestas e campos – 1872/75. Curitiba: Imprensa Oficial do Paraná, 1878.
- BRASIL. Instrução Normativa no 1 de 25 de março de 2015. Diário Oficial [da] República Federativa do Brasil. Brasília: 26 março 2015, Seção 1, p. 11, 2015.
- CLARKE, David L. Models in Archaeology. London: Routledge, 1972.
- CONOLLY, James; LAKE, Mark. Geographical information systems in archaeology. Cambridge: Cambridge University Press, 2006.
- COPPOCK, J. Terry; RHIND, David W. The history of GIS. IN: LONGLEY, Paul; GOODCHILD, Michael F.; MAGUIRE, David J.; RHIND, David W. Geographical Information Systems: principles and applications. New York: Longman, Harlow/Wiley, 1991, v. 2. p. 21–43.
- D’AVILA, Luísa Nunes. Caminhos e lugares: modelos de mobilidade e sistema de assentamento entre os proto-Jê meridionais na região de Campo Belo do Sul, SC. Dissertação (Mestrado em Antropologia), Universidade Federal de Pelotas, Rio Grande do Sul, 2017.
- EMBRAPA/EMATER. Mapa de Solos do Paraná. Disponível em: www.itcg.pr.gov.br/modules/faq/category.php?categoryid=9#. Acessado em 01/09/2019.
- FERNANDES, José Loureiro. A dying people. Bulletin of the International Committee on Urgent Anthropological Research, n. 2, p. 22–26, 1959a.
- FERNANDES, José Loureiro. Os índios da Serra dos Dourados. Separata da III Reunião Brasileira de Antropologia, p. 27–46, 1959b.
- FERNANDES, José Loureiro. Os índios da Serra dos Dourados. Bulletin of the International Committee on Urgent Anthropological and Ethnological Research, n. 5, p. 151–154, 1962.
- FONSECA, João Aires A. Da. Padrões de distribuição espacial e modelos preditivos: os sítios arqueológicos no baixo curso dos rios Nhamundá e Trombetas. Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, v. 13, n. 2, p. 353–376, 2018.
- FONSECA, João Aires Ataíde Da. Levantamento regional na arqueologia amazônica: O uso de sistema de informação geográfica e sensoriamento remoto. Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, v. 8, n. 3, p. 675–690, 2013.
- GAMBLE, Clive. Archaeology, geography and time. Progress in Human Geography, v. 11, n. 2, p. 227–246, 1987.

- GILLINGS, Mark; WHEATLEY, David. Geographic Information Systems. IN: MASCHNER, Herbert D.G.; CHIPPINDALE, Christopher. Handbook of Archaeological Methods. Walnut Creek, CA: Altamira Press, 2005. v. 1, p. 373–422.
- HODDER, Ian; ORTON, Clive. Spatial Analysis in Archaeology. Cambridge: Cambridge University Press, 1976.
- HOLTON, Jerry Thomas. Predictive Model of Archaeological Sites of the Hopi Reservation of Northeastern Arizona. Master Thesis (MA in Science in Geographic Information Systems), University of Redlands, 2014.
- HU, Di. Advancing Theory? Landscape Archaeology and Geographical Information Systems. Papers from the Institute of Archaeology, v. 21, p. 80, 2012.
- IPARDES/ITCG. Hidrografia do Estado do Paraná. Material Vetorizado com base nas Cartas Topográficas na escala de 1:250.000. Disponível em: <http://www.itcg.pr.gov.br/modules/faq/category.php?categoryid=9#>. Acessado em 20/09/2019.
- JAMOVI, Project. Jamovi - Version 1.1.5. Disponível em: <<https://www.jamovi.org>>.
- KELLER, José; KELLER, Francisco. Exploração do rio Ivai. Extrato de relatório apresentado à assembléia legislativa da Província do Paraná, na 1a sessão da 8a legislatura pelo Presidente Dr. André Augusto de Padua Fleury, 1866.
- KIPNIS, Renato. O uso de modelos preditivos para diagnosticar recursos arqueológicos em áreas a serem afetadas por empreendimentos de impacto ambiental. IN: CALDARELLI, Solange Bezerra. Atas do Simpósio sobre Política Nacional do Meio Ambiente e Patrimônio Cultural. Goiânia: Instituto Goiano de Pré-História e Antropologia, 1997, p. 34–51.
- KOHLER, Timothy A.; PARKER, Sandra C. Predictive Models for Archaeological Resource Location. Advances in Archaeological Method and Theory, v. 9, p. 397–452, 1986.
- KOZÁK, Vladimir; BAXTER, David; WILLIAMSON, Laila; CARNEIRO, Robert L. Os índios Héta: peixe em lagoa seca. Boletim do Instituto Histórico, Geográfico e Etnográfico Paranaense, v. 38, 1981.
- KOZÁK, Vladimir; BAXTER, David; WILLIAMSON, Laila; CARNEIRO, Robert L. The Héta Indians: fish in a dry pond. Anthropological Papers of the American Museum of Natural History, v. 55, n. 6, 1979.
- KOZLOWSKI, Henrique de Sena. Modelagem preditiva e ocupação Jê meridional na encosta catarinense. Dissertação (Mestrado em Arqueologia), Universidade de São Paulo, 2018.
- KOZLOWSKI, Henrique de Sena; CORTELETTI, Rafael. Paisagem de mobilidade Jê Meridional do Alto Vale do Ribeira de Iguape Resumo. Arqueologia, v. 22, n. 1, p. xx–xx, 2019.
- KOZLOWSKI, Henrique de Sena; DEBLASIS, Paulo. Perspectivas de modelagem preditiva arqueológica:

estudo de caso na Bacia do Rio Tubarão – Santa Catarina. *Especiaria - Caderno de Ciências Humanas*, v. 18, n. 33, p. 72–92, 2019.

KVAMME, Kenneth L. Development and testing of quantitative models. IN: JUDGE, W. James; SEBASTIAN, Lynne. *Quantifying the present and predicting the past: theory, method, and application of archaeological predictive modeling*. Denver, Co: U.S. Department of the Interior, Bureau of Land Management Service Center, 1988, p. 325–428.

LAMING-EMPERAIRE, Annette; MENEZES, Maria José; ANDREATA, Margarida Davina. O Trabalho da pedra entre os Xetá da Serra dos Dourados, Estado do Paraná. *Coleção Museu Paulista: série ensaios*, v. 2, p. 19–82, 1978.

MAACK/ITCG/IBGE. Mapa Fitogeográfico do Paraná 1989/1990. Disponível em: <http://www.itcg.pr.gov.br/modules/faq/category.php?categoryid=9#>. Acessado em 01/09/2019.

MILHEIRA, Rafael Guedes; DE SOUZA, Jonas Gregorio; IRIARTE, José. Water, movement and landscape ordering: A GIS-based analysis for understanding the mobility system of late Holocene mound-builders in southern Brazil. *Journal of Archaeological Science*, v. 111, p. 105014, 2019.

MOTA, Lúcio Tadeu. *Os Xetá no vale do rio Ivaí 1840 - 1920*. Maringá: EDUEM, 2013.

NASA/JPL. NASA Shuttle Radar Topography Mission Global 1 arc second. Disponível em: <https://doi.org/10.5067/MEaSURES/SRTM/SRTMGL1.003>. Acessado em: 20/09/2019.

PEREZ, Glauco Constantino. *Arqueologia paulista e o marcador cerâmico como delimitador de fronteira étnica: um estudo das regiões sul e oeste do estado de São Paulo*. Tese (Doutorado em Arqueologia), Universidade de São Paulo, 2018.

PEREZ, Glauco Constantino; COUTINHO AFONSO, Marisa; MOTA, Lúcio Tadeu. Métodos de análise espacial para sítios arqueológicos: um modelo preditivo para o Estado de São Paulo. *Cadernos do LEPAARQ (UFPEL)*, v. 15, n. 30, p. 98, 2018.

RIRIS, Philip George Constantine. *Predictive modelling in Misiones Province, Argentina using GIS: a case study of the Taquara/Itararé tradition*. 104 f. Unpublished dissertation (MA in Archaeology), University of Exeter, 2010.

RODRIGUES, Aryon Dall'Igna. A língua dos índios Xetá como dialeto Guaraní. *Cadernos de Estudos Lingüísticos* 1, v. 1, p. 7–11, 1978.

RODRIGUES, Aryon Dall'Igna. Relações internas na família lingüística Tupí-Guaraní. *Revista de Antropologia*, v. 27/28, p. 33–53, 1985.

ROOD, Ronald J. Spatial analysis in archaeology: Historical developments and modern applications. *Lambda Alpha Journal of Man*, v. 14, p. 25–60, 1982.

ROSA, Roberto. Geotecnologias na Geografia aplicada. *Revista do Departamento de Geografia*, v. 16, p. 81–90, 2005.

- SILVA, Carmen Lucia. Em busca da sociedade perdida: o trabalho da memória Xetá. Tese (Doutorado em Antropologia Social), Universidade de Brasília, Brasília, 2003.
- SILVA, Carmen Lucia. Sobreviventes do extermínio: uma etnografia das narrativas e lembranças da sociedade Xetá. Dissertação (Mestrado em Antropologia Social), Universidade Federal de Santa Catarina, Florianópolis, 1998.
- SILVA, Bruno Labrador Rodrigues Da. Sistema de assentamento proto-Jê meridional no alto rio Canoas. Dissertação (Mestrado em Arqueologia), Universidade de São Paulo, São Paulo, 2018.
- TSUCHIYA, Ítalo; TACHIBANA, Vilma Mayumi; IMAI, Nilton Nobuhiro; THOMAZ, Rosângela Custódio Cortez. Utilização de regressão logística para avaliação de prováveis áreas de ocupações pretéritas. Anais do Simpósio Brasileiro de Geomática, Presidente Prudente -SP, 9-13 de julho de 2002, p. 565–573, 2002.
- VERHAGEN, Philip. Case Studies in Archaeological Predictive Modelling. Leiden: Leiden University Press, 2007.
- VERHAGEN, Philip; NUNINGER, Laure; TOURNEUX, François-Pierre; BERTONCELLO, Frédérique; JENESON, Karen. Introducing the Human Factor in Predictive Modelling: a work in progress. IN: EARL, Graeme; SLY, Tim; CHRYSANTHI, Angeliki; MURRIETA-FLORES, Patricia; PAPADOPOULOS, Constantinos; ROMANOWSKA, Iza; WHEATLEY, David. Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012. Amsterdam: Amsterdam University Press, 2013, p. 379–388.
- VERHAGEN, Philip; WHITLEY, Thomas G. Integrating Archaeological Theory and Predictive Modeling: A Live Report from the Scene. *Journal of Archaeological Method and Theory*, v. 19, n. 1, p. 49–100, 2012.
- WATSON, Patty Jo; LEBLANC, Steven A.; REDMAN, Charles L. Explanation in archaeology: an explicitly scientific approach. New York: Columbia University Press, 1971.
- WHEATLEY, David. Making space for an archaeology of place. *Internet Archaeology*, v. 15, n. 15, 2004.
- WHEATLEY, David; GILLINGS, Mark. Spatial Technology and Archaeology: The archaeological applications of GIS. New York: Taylor & Francis, 2002.
- WILLEY, Gordon R. Prehistoric settlement patterns in the Virú Valley, Perú. Washington, DC: Smithsonian Institution, 1953, v. 155.

Material suplementar 1: Valores obtidos de cada variável independente (elevação, declividade, custo distância água, distância água, solo, vegetação) para a variável

ID	X	Y	Dependente	Elevação	Declividade	Custo distância água	Distância água	Solo	Vegetação
Guayará	254537	7406262	1	300	1,358170033	1780,369995	457,9299927	1	3
sn1	250403	7411462	1	281	1,440520048	129,0579987	150	2	3
sn2	249914	7410809	1	297	6,565070152	1224,719971	540	2	3
sn3	249749	7408764	1	290	3,866230011	1611,219971	313,2090149	2	3
1	260297	7413136	1	283	3,200500011	658,4140015	228,4730072	1	5
2	256485	7392901	1	349	3,392090082	0	0	1	5
3	257448	7393131	1	343	8,488409996	791,8220215	120	2	3
4	258561	7393770	1	334	2,584609985	948,5269775	150	2	3
5	259785	7392993	1	356	6,793970108	254,9850006	30	1	3
6	260187	7391187	1	368	4,287690163	903,8319702	240	2	3
7	258868	7391216	1	349	1,979439974	1142,839966	301,4960022	1	3
8	258962	7390498	1	349	3,408970118	978,5579834	152,970993	1	3
9	255411	7392565	1	370	8,696789742	602,03302	123,6930008	1	5
10	255897	7393728	1	380	4,096479893	3264,679932	997,2459717	1	5
11	256993	7394848	1	376	5,158740044	4789,620117	1279,140015	2	5
12	256478	7396357	1	377	5,180809975	1194,459961	241,8679962	1	5
13	257509	7398814	1	316	2,584609985	1273,910034	692,6040039	2	5
14	256767	7400788	1	325	1,697530031	731,5289917	216,3329926	2	3
15	255256	7404519	1	309	4,791550159	2467,459961	725,6030273	1	3
16	250170	7409215	1	297	4,791550159	299,6839905	60	2	3
18	261637	7410379	1	294	0,96046102	289,446991	210	2	5
1	242091	7427349	0	260	2,918869972	1535,670044	657,9509888	2	5
2	264912	7414119	0	395	2,375880003	2218,969971	931,4509888	2	5
3	265789	7410346	0	364	7,406610012	6078,990234	1467,550049	2	5
4	239789	7416994	0	323	3,791369915	394,381012	192,0939941	1	5
5	243905	7415721	0	334	3,791369915	3876,76001	1140,390015	2	5

6	264173	7389296	0	406	1,920379996	243,4129944	150	1	5
7	257308	7412657	0	325	6,642280102	1894,800049	360	1	5
8	245252	7429095	0	278	3,791369915	2754,52002	834,0859985	2	3
9	260096	7382559	0	440	6,237269878	176,572998	30	1	5
10	261687	7399211	0	412	0,96046102	5547,180176	1816,180054	2	5
11	246334	7403676	0	390	2,918869972	4819,129883	1481,890015	1	5
12	250239	7397622	0	347	2,714809895	1500,280029	442,9450073	1	3
13	250904	7392512	0	393	1,073799968	2605,540039	780	2	5
14	257118	7420556	0	285	4,885990143	1954,369995	713,09198	1	3
15	258654	7407969	0	397	6,760570049	870,9240112	212,1320038	2	5
16	245899	7408003	0	379	5,511320114	3625,439941	1087,060059	1	5
17	262874	7386297	0	421	6,219039917	3603,810059	1207,849976	1	5
18	263333	7403468	0	401	5,511320114	2822,219971	844,8079834	2	5
19	249142	7418673	0	309	1,22428	3736,26001	1722,849976	2	5
20	251292	7404748	0	380	3,954159975	4420,669922	1879,48999	2	5
21	257185	7426055	0	248	1,440520048	426,29599	402,4920044	2	3