

Perspectivas de modelagem preditiva arqueológica: estudo de caso na Bacia do Rio Tubarão – Santa Catarina

Henrique de Sena Kozlowski

Mestre em Arqueologia pela Universidade de São Paulo, USP.

E-mail: kozloeski.henrique@gmail.com

Paulo DeBlasis

Doutorado e Mestrado pelo Museu de Arqueologia e Etnologia da Universidade de São Paulo (MAE/USP)

E-mail: deblasis@usp.br

Recebido em: 30/12/2018.

Aprovado em: 28/03/2019.

Resumo: A modelagem preditiva arqueológica é uma ferramenta de resolução pragmática de problemas relacionados com a localização de sítios arqueológicos, relacionada principalmente com a diminuição de custos e aumento na eficiência de atividades de *survey*. Há uma preponderância na utilização de dois métodos distintos de modelagem preditiva: uma abordagem chamada de dedutiva e outra, indutiva. Neste artigo, iremos discutir aspectos destes dois tipos de técnicas de modelagem preditiva, aplicando os dois métodos mais recorrentes na literatura em uma área piloto na região sul de Santa Catarina, a bacia do Rio Capivaras. Os modelos preditivos foram utilizados para avaliar a distribuição espacial de um tipo particular de sítio na região, caracterizado pela presença de uma mancha circular de solo bastante escuro com materiais líticos e resíduos de atividades de combustão. Com objetivo de avaliar criticamente os diferentes métodos de modelagem, foram realizados alguns testes comparativos, demonstrando pontos positivos e negativos de cada método.

Palavras-chave: Modelagem preditiva arqueológica. Sítios arqueológicos.

Prospects of archaeological predictive modeling: case study in the Tubarão River Basin - Santa Catarina

Abstract: Archaeological predictive modelling is a methodological tool for evaluating archaeological site location and distribution patterns. Its use is deeply rooted in cost and efficiency management in survey activities. There are two main methodological approaches in predictive modelling: deductive and inductive. This paper discusses aspects of these two techniques, applying both methods in a pilot area at the southern region of Santa Catarina State in Brazil. Predictive models were used to test spatial distribution of a specific kind of site in the region, small circular dark soil locations with lithic materials and residues of combustion activities. Critically assessing these different methods of predictive modelling, comparative tests were applied, exploring positive and negative points of each method.

Keywords: Archeological predictive modeling. Archeological sites.

Introdução

Desde a metade do século XX, a literatura arqueológica tem se preocupado com o desenvolvimento de modelos capazes de articular sítios arqueológicos em uma abordagem espacial (WILLEY, 1953; HODDER; ORTON, 1976; PLOG; HILL 1971). Aliados à perspectiva espacial, surgem também interesses relacionados com o desenvolvimento de ferramentas e métodos de identificação de sítios arqueológicos através da análise de características do espaço, tendo em vista o cenário teórico vigente na época, onde estudos quantitativos ganharam grande destaque, influenciados principalmente pelas correntes da *New Geography* (TILLEY, 1994). Esse interesse culminou na criação da modelagem preditiva arqueológica, um esforço na previsão da presença ou ausência de sítios arqueológicos em áreas determinadas (KOHLENER; PARKER, 1986, p. 400).

Apesar de ter surgido em um contexto processualista, a modelagem preditiva arqueológica não busca criar leis gerais do comportamento humano relacionadas à localização de sítios arqueológicos, e sim resolver de forma pragmática problemas relacionados com a identificação de sítios em áreas pouco exploradas. O aspecto pragmático da modelagem preditiva aliado à necessidade de se conhecer sítios e áreas de interesse arqueológico para fins de pesquisa científica e gestão culminaram na obra seminal de Judge e Sebastian (1988), publicada pelo Departamento de Interior dos EUA (DOI). A modelagem preditiva surge como método de relativo baixo custo que aumenta a eficiência de *surveys* dentro de construções logísticas e orçamentárias tão comuns à pesquisa arqueológica (VERHAGEN, 2007).

As discussões de modelagem preditiva são amplificadas conforme a crescente facilidade de acesso a *softwares* de Sistemas de Informação Geográfica (SIG) e também devido ao barateamento de computadores capazes de processar grandes volumes de dados espaciais a partir da década de 1990 e início dos anos 2000. Esse crescimento é exemplificado pela publicação de importantes trabalhos na literatura internacional (KOHLENER; PARKER, 1986; KVAMME, 1990; WESCOTT; BRANDON, 2000; WHEATLEY; GILLINGS, 2002) e nacional (KASHIMOTO, 1997; KIPNIS, 1997; KNEIP, 2004; THOMAZ, 2002; TSUCHIYA et al., 2002). A modelagem preditiva pode ser especialmente útil para a arqueologia de contrato focada em projetos de licenciamento ambiental que, na maioria das vezes, apresentam um cronograma restrito. Contudo, a modelagem preditiva também pode auxiliar pesquisas acadêmicas como apresentado, por exemplo, em Fonseca Júnior (2013).

Neste artigo, iremos apresentar os métodos e resultados das atividades de modelagem preditiva na microbacia do Rio Capivaras, na área conhecida como encosta catarinense

(Figura 1), decorrentes do projeto *Modelagem Preditiva e Ocupação Jê Meridional na Encosta Catarinense* (KOZLOWSKI, 2018) desenvolvido no âmbito do projeto *Jê Landscapes of Southern Brazil*. A bacia do Rio Capivaras está localizada no município de São Martinho, próxima às encostas da Serra do Tabuleiro, na porção nordeste da Bacia do Rio Tubarão. Trata-se de uma área de 69 km² com declives acentuados e vegetação alterada pelas atividades agrícolas (predominantemente o reflorestamento de pinus e criação de gado).

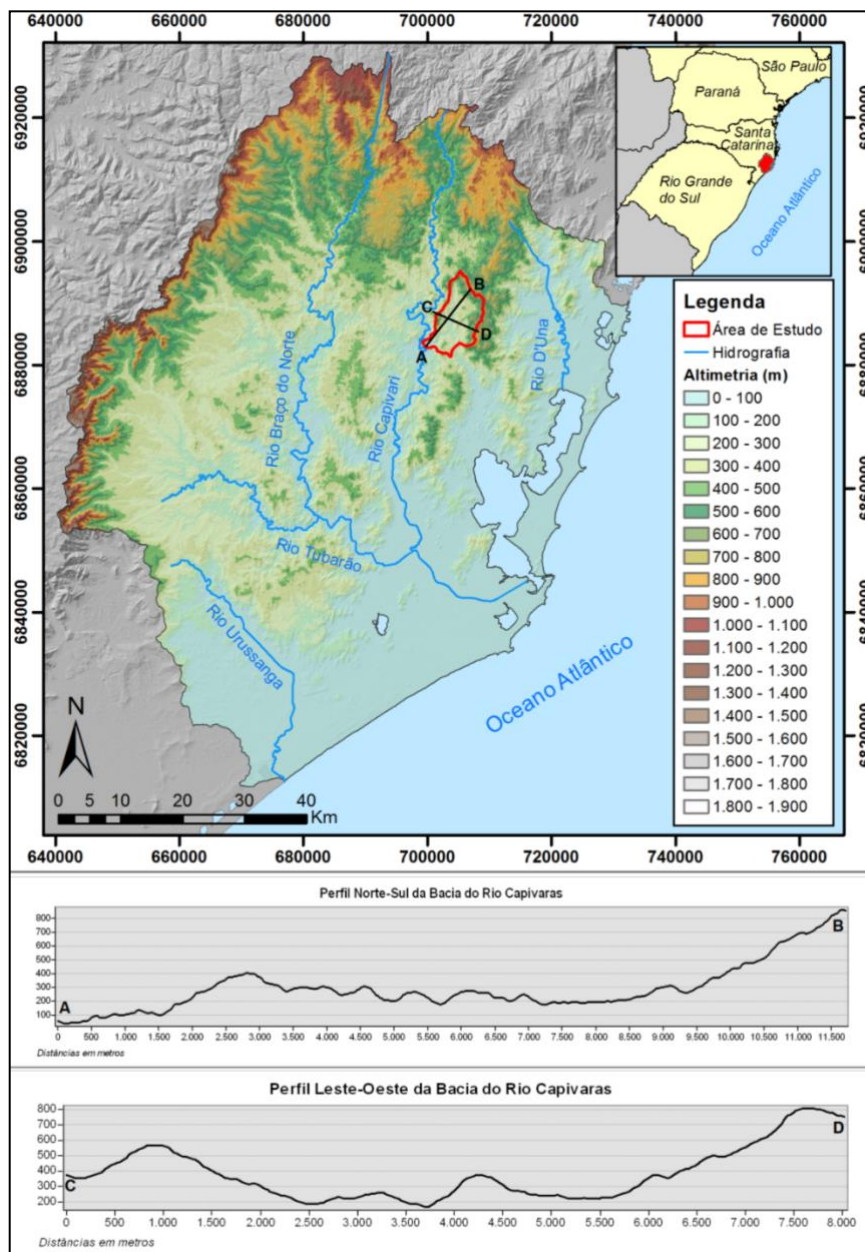


Figura 1 - Mapa de localização da área de estudo da bacia do Rio Capivaras (base: GIS Grupep-UNISUL).

A área de estudo faz parte de uma zona ecológica de transição entre os ambientes do litoral e do planalto catarinense, onde predomina o bioma da Mata Atlântica. Esta região está

inserida no contexto do território Jê Meridional e o estudo dos sítios nesta paisagem pode auxiliar a compreender a articulação da ocupação Jê do Sul no litoral e planalto, principalmente a partir da compreensão dos principais rios da região (Urussanga, Tubarão, Braço do Norte e Capivari) como possíveis rotas de conexão entre estas diferentes paisagens.

Nome do Sítio	Sigla	UTM SIRGAS 2000 Zona 22S		Tipo de Sítio	Material
		X	Y		
Vargem do Cedro 01	VC-01	704141	6887961	Mancha escura	
Vargem do Cedro 02	VC-02	704236	6887906	Mancha escura	Lítico
Vargem do Cedro 03	VC-03	704358	6887827	Mancha escura	
Vargem do Cedro 04	VC-04	704387	6887568	Mancha escura	
Vargem do Cedro 05	VC-05	704403	6893929	Mancha escura	Lítico
Vargem do Cedro 06	VC-06	704396	6890898	Mancha escura	
Vargem do Cedro 07	VC-07	704467	6889295	Mancha escura	
Vargem do Cedro 08	VC-08	705269	6885431	Mancha escura	Cerâmico e Lítico
Vargem do Cedro 09	VC-09	702822	6884827	Mancha escura	
Vargem do Cedro 10	VC-10	703859	6885581	Mancha escura	
Vargem do Cedro 11	VC-11	702251	6884930	Mancha escura	
Vargem do Cedro 12	VC-12	704443	6891465	Mancha escura	Lítico
Vargem do Cedro 13	VC-13	704305	6889298	Mancha escura	
Vargem do Cedro 14	VC-14	704315	6890827	Mancha escura	
Vargem do Cedro 15	VC-15	704217	6887010	Mancha escura	
Vargem do Cedro 16	VC-16	706370	6888491	Lítico	Lítico
Vargem do Cedro 17	VC-17	706668	6890280	Mancha escura	
Vargem do Cedro 18	VC-18	704563	6892225	Mancha escura	
Vargem do Cedro 19	VC-19	702986	6889506	Lítico	Lítico
Vargem do Cedro 20	VC-20	704305	6890130	Mancha escura	
Vargem do Cedro 21	VC-21	704267	6890174	Mancha escura	
Vargem do Cedro 22	VC-22	704397	6891375	Mancha escura	
Vargem do Cedro 23	VC-23	704723	6893546	Lítico	Lítico

Vargem do Cedro 24	VC-24	704316	6890076	Mancha escura	
Vargem do Cedro 25	VC-25	703924	6885188	Mancha escura	
Vargem do Cedro 26	VC-26	704406	6884095	Mancha escura	
Vargem do Cedro 27	VC-27	704429	6883889	Mancha escura	
Canto dos Hoepers 1	CH-01	705508	6890042	Mancha escura	
Canto dos Hoepers 2	CH-02	704715	6889580	Mancha escura	
Canto dos Hoepers 3	CH-03	706131	6890350	Mancha escura	
Canto dos Hoepers 4	CH-04	705756	6890257	Lítico	Lítico
Canto dos Hoepers 5	CH-05	707402	6889490	Mancha escura	
Canto dos Hoepers 6	CH-06	706988	6889396	Mancha escura	
São Luís 1	SL-01	703905	6881296	Lítico	Lítico
São Luís 2	SL-01	703346	6880519	Lítico	Lítico
Santa Maria 1	SM-01	701675	6895899	Mancha escura	

Tabela 1 - Lista e classificação de sítios identificados na área de pesquisa. Fonte: Kozłowski (2018)

Entretanto, o registro arqueológico relacionado a grupos Jê do Sul nesta região do estado ainda é pouco conhecido. A maior quantidade de pesquisas feitas na bacia do rio Tubarão é decorrente de atividades realizadas pelo Grupo de Pesquisa em Educação Patrimonial e Arqueologia da Unisul (GRUPEP-Unisul) e também trabalhos de arqueologia de contrato. A área de pesquisa foi visitada na década de 1970 para a construção do Parque Estadual da Serra do Tabuleiro (EBLE; REIS, 1976), porém o banco de dados mais atualizado (FARIAS; KNEIP, 2010) não continha as coordenadas geográficas dos oito sítios localizados naquele trabalho. Assim, foi realizada uma prospecção durante um mês quando foi possível, através de estratégias oportunísticas e sistemáticas, a fim de identificar 36 sítios arqueológicos (Tabela 1), incluindo três dos oito sítios registrados por Eble e Reis (1976). Estes sítios foram classificados em dois tipos distintos: sítios líticos à céu aberto e sítios com manchas de solo escuro. Estes tipos de sítio são bastante comuns na região da encosta catarinense e os líticos, especialmente, são associados a horizontes tardios da ocupação da encosta por grupos caçadores-coletores tradicionalmente identificados pela tecnologia lítica Umbu. Entender as relações entre estes tipos de sítio e, ademais, entender como se conectam com a expansão dos grupos Jê do Sul, são os maiores desafios para a pesquisa arqueológica nesta região meridional da encosta catarinense, motivando este estudo.

Modelando a Bacia do Rio Capivaras

Em linhas gerais, existem dois tipos de abordagem de construção de modelos preditivos: abordagem dedutiva ou *theory-driven* e a abordagem indutiva ou *data-driven* (WHEATLEY; GILLINGS, 2002, p. 149). Com o resultado das atividades de prospecção, foi possível criar uma base de dados geográficos para elaboração de modelos preditivos utilizando as técnicas citadas, a fim de comparar suas vantagens e desvantagens e explorar suas particularidades.

A abordagem dedutiva da modelagem preditiva é caracterizada como forma de modelagem orientada pelas experiências prévias do arqueólogo com áreas preferenciais de localização de sítios arqueológicos. Esta abordagem é colocada em oposição à modelagem de abordagem indutiva (*data-driven*) que, por sua vez, seria orientada apenas por dados empíricos tratados de forma estatística para obtenção de padrões (WHEATLEY; GILLINGS, 2002). Todavia, é possível perceber que nesta discussão ambas as abordagens tratam com teoria e dados empíricos, pois estes são elementos mutuamente dependentes.

Para a elaboração dos modelos, foram selecionadas quatro variáveis ambientais que correspondem a fatores que integram o processo de escolha de locais para ocupação: altitude (elevação do terreno em relação ao nível do mar), declividade, incidência de radiação solar e distância da água. Esta escolha, arbitrária, baseia-se em seu uso recorrente na literatura (RIRIS, 2010; FONSECA JÚNIOR, 2013; VERHAGEN, 2007), pois correspondem a características físicas da paisagem que influenciam a escolha de lugares para ocupação. Variáveis culturais, ainda pouco conhecidas na área, não foram consideradas neste estudo.

Para representar os diferentes compartimentos do intenso relevo da região, a área foi estratificada em compartimentos (fundo de vale, encostas e topos de morro) a partir de cotas altimétricas de referência (300 e 600 metros). A variável declividade do relevo ajuda a definir as áreas de melhor acesso e menor gasto energético de mobilidade, pois uma área de relevo mais declivoso gera respostas diferentes de movimentação do que uma área de relevo mais suave. Araújo (2001), ao comentar sobre a localização dos sítios no vale do rio Paranapanema, em São Paulo, apresenta uma série de sítios localizados na face norte das vertentes, que recebe maior incidência de radiação solar ao longo do ano, argumentando que a variável radiação solar ajuda a identificar as áreas de maior conforto térmico na paisagem. Por fim, distância da água é uma das variáveis que apresenta grande potencial para a identificação de locais preferenciais, já que o acesso à água está relacionado com hidratação, alimentação e mobilidade. Esta variável não reflete distâncias euclidianas até as fontes de água, e sim a

facilidade de acesso à fonte baseando-se na declividade do terreno. A obtenção dos dados de elevação foi feita através de dados disponíveis *online* no *EarthExplorer*, serviço de imagens de satélite e modelos digitais produzidos pelo USGS (*United States Geological Survey*). Os dados secundários (declividade e radiação solar) e terciário (distância da água), foram produzidos através da aplicação de algoritmos no *software* ArcGIS 10.1. A Tabela 2 apresenta os metadados das variáveis utilizadas.

Tabela 2 - Metadados das variáveis. Fonte: Kozłowski (2018)

Variável	Sigla	Classe de dados	Descrição	Fonte
Presença de sítios (Variável Dependente)	sitiosjev1 pontosaleatoriosv1	Primário	Arquivos vetoriais <i>shapefile</i> de pontos aleatórios e sítios arqueológicos previamente identificados para compor a variável dependente.	GRUPEP- Unisul
Elevação (m)	elevacaorh9	Primário	Arquivo <i>raster</i> do modelo digital de elevação ASTER. Resolução espacial de 30m.	ASTER GDEM METI NASA
Declividade (°)	declivgraus	Secundário	Arquivo <i>raster</i> produzido a partir do modelo digital de elevação com a ferramenta <i>Slope</i>	-
Distância da água	distagua	Terciário	Arquivo <i>raster</i> produzido a partir da declividade, com base na ferramenta <i>Cost Distance</i> .	-
Radiação Solar (WH/m²)	radsolrh9	Secundário	Arquivo <i>raster</i> produzido a partir do modelo digital de elevação com a ferramenta <i>Area Solar Radiation</i>	-

Modelagem preditiva dedutiva

A abordagem dedutiva na modelagem preditiva consiste na construção de modelos com base em pressupostos *a priori* do arqueólogo sobre características espaciais recorrentes na localização de sítios arqueológicos. Ainda que a abordagem indutiva também apresente um momento de escolha arbitrária de características espaciais na forma de variáveis espaciais avaliadas por um modelo, na abordagem dedutiva este aspecto arbitrário permeia todas as etapas de modelagem.

Um método recorrente para a produção de modelos de abordagem dedutiva é apresentado por Wheatley e Gillings (2002, p. 151), baseado em regras de álgebra de mapas através de uma função matemática de intersecção e soma ponderada de diferentes planos de informação. O resultado da aplicação desta função matemática é a definição de áreas que concentram mais características favoráveis do que outras. A resposta pode então ser exibida

como respostas binárias de presença-ausência ou então de forma qualitativa, definindo áreas de alta, média ou baixa probabilidade. No presente artigo, a aplicação da abordagem dedutiva teve como objetivo produzir um modelo que apresentasse o resultado de forma qualitativa.

Utilizando as variáveis de elevação, declividade, incidência de radiação solar e distância da água definidas como guia para os modelos, aplicamos a classificação de valores de acordo com pesos de importância arbitrários, definidos a partir de dados conhecidos da região da encosta a partir de Farias *et al.* (2018) e ainda não testados para o vale do rio Capivaras. Para a variável *elevação* foram definidos três estratos que representam diferentes compartimentos da paisagem: 0 a 300 metros, que corresponde aos fundos de vale; 300 a 600 metros, que corresponde às encostas e elevações médias; e 600 a 900 metros, que corresponde aos topos da Serra do Tabuleiro e suas ramificações. Para a variável *declividade* foram definidas cinco classes: 0° a 15°, 15° a 30°, 30° a 45°, 45° a 60° e 60° a 75°. A variável *distância da água* foi dividida em três classes, representando locais de distância próxima, média distância e longa distância. Por fim, a variável *radiação solar* foi dividida também em três categorias: alta, média e baixa incidência de radiação. Utilizando a ferramenta *Reclassify*, foram atribuídos pesos distintos para cada classe de variáveis. A Tabela 3 apresenta as classes de variáveis e seus respectivos pesos.

Os compartimentos de fundo de vale foram privilegiados, portanto receberam maior peso quando comparados com os compartimentos mais altos, pois a maioria dos sítios na região encontram-se próximos aos maiores corpos d'água. As declividades até 15° são as mais importantes no cálculo do modelo, pois correspondem às áreas menos acentuadas da paisagem e, portanto, de melhor mobilidade que áreas acima de 30°. As variáveis restantes tiveram os maiores pesos definidos de acordo com a menor distância da água e a maior incidência de radiação solar, quanto maior a distância e menor a incidência de radiação solar, menor o peso atribuído. Em seguida, foram atribuídos também pesos relativos para as variáveis analisadas. Elevação e declividade foram selecionadas como mais importantes na localização dos sítios na bacia do Rio Capivaras, recebendo o peso 3, tendo em vista que a maior parte dos sítios na região da encosta se localiza nos fundos de vale e áreas de declividade suave. À radiação solar foi atribuído peso 2, pois está relacionada ao conforto térmico e iluminação, podendo ser bastante relevante na escolha do lugar de ocupação, já que a região da encosta apresenta temperaturas amenas e frias no inverno. A variável distância da água recebeu peso 1, tendo em vista que a região apresenta grande quantidade de recursos hídricos e falta de água não seria problema em praticamente qualquer lugar da área de pesquisa. Assim, foi produzida uma fórmula final aplicada no *software*:

$$\text{Probabilidade} = \text{Elevação} \times 3 + \text{Declividade} \times 3 + \text{Radiação Solar} \times 2 + \text{Distância da Água} \times 1$$

Esta equação auxilia a identificação das áreas que apresentam as regiões de fundo de vale e baixa vertente, com declividade de até 15°, localizadas na face norte das vertentes e próximas a fontes de água, que correspondem aos locais mais propícios para a identificação de sítios arqueológicos. No resultado final foi feita uma nova classificação, estabelecendo classes qualitativas de alta, média e baixa probabilidade de ocorrência de sítios arqueológicos (Figura 2).

Tabela 3 - Reclassificação de valores das variáveis. Fonte: Kozłowski (2018).

Variável	Valor Original	Peso atribuído
Elevação (m)	0 - 300	3
	300 - 600	2
	600 - 900	1
Declividade (°)	0° - 15°	4
	15° - 30°	2
	30° - 45°	1
	45° - 60°	0
	60° - 75°	0
Distância da Água	0 - 5398,78	3
	5398,78 - 15116,59	2
	15116,59 - 55067	1
Radiação Solar (WH/m ²)	467.289,06 - 1.348.608,79	3
	1.348.608,79 - 1.509.352,64	2
	1.509.352,64 - 1.880.730	1

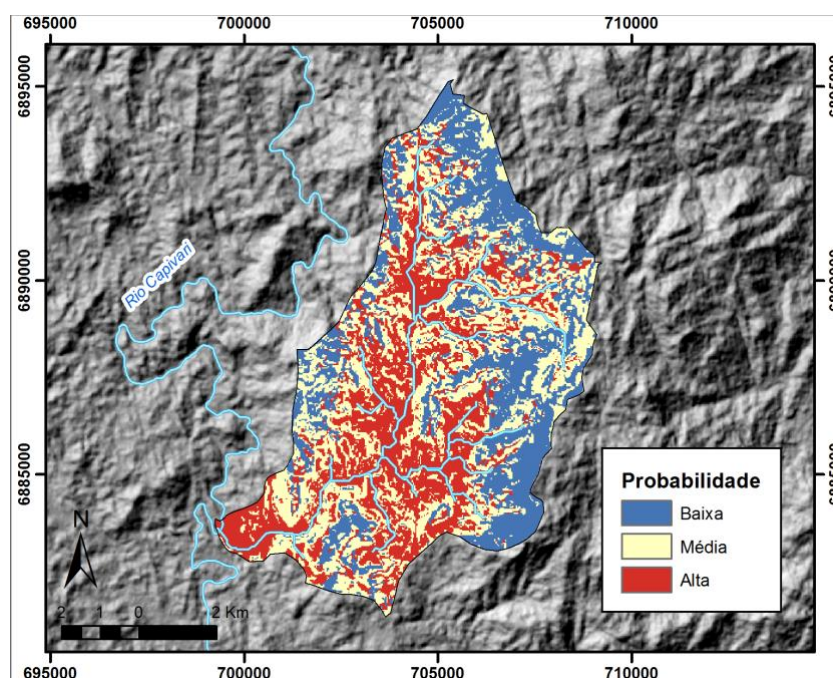


Figura 2 - Modelo preditivo de abordagem dedutiva. Fonte: Kozłowski (2018).

Modelagem preditiva indutiva

A abordagem indutiva da modelagem preditiva, de acordo com Wheatley e Gillings (2002, p. 149), é uma forma de obter padrões de distribuição de sítios em uma região através da utilização de características de um conjunto de dados previamente conhecidos utilizando ferramentas estatísticas. Este método de modelagem parte do princípio de que os dados “falam” por si mesmos, ou seja, o autor do modelo não deve interferir na definição da probabilidade de ocorrência de sítios arqueológicos na área de estudo. Esta abordagem também é conhecida como *data-driven*, ou seja, orientada pelos dados e não por uma teoria externa. Contudo, como já discutido anteriormente, este método também apresenta uma etapa teórica importante, pois o autor do modelo precisa definir o conjunto de variáveis a ser analisado, e esta definição parte de conhecimentos teóricos apriorísticos do indivíduo.

O modelo de abordagem indutiva apresenta duas importantes demandas: a existência de um conjunto de dados geográficos de localização de sítios e a definição de um método estatístico de análise. Para o modelo apresentado neste artigo, foi utilizado o banco de dados de sítios identificados após a etapa de prospecção na área da bacia do Rio Capivaras e o método estatístico aplicado foi a regressão logística, que é um dos mais utilizados na literatura (JUDGE; SEBASTIAN, 1988; KOHLER; PARKER, 1986; MERENCIO, 2013; RIRIS, 2010; THOMAZ, 2002; TSUCHIYA ET AL., 2002; VERHAGEN, 2007; WESCOTT; BRANDON, 2000). A função da regressão logística é capaz de avaliar a capacidade explanatória de cada variável para a distribuição espacial do conjunto de sítios conhecidos podendo projetar a probabilidade em uma área maior que a da amostra inicial (RIRIS, 2010).

A abordagem indutiva não necessariamente aponta sítios ainda não descobertos, mas, sim, mapeia locais com características similares às aquelas amostradas inicialmente (KVAMME, 1988). O conjunto de dados geográficos necessários para a elaboração dos modelos é composto por dois elementos: os sítios conhecidos e um mesmo número de locais aleatórios onde não há presença de sítios arqueológicos, pois a regressão logística opera com valores binários de certeza ou incerteza da presença de sítios.

Em ambiente SIG, foram realizadas operações para amostrar os valores das variáveis elevação, declividade, distância da água e radiação solar de 28 sítios identificados na prospecção (Tabela 1) e 28 pontos aleatórios de onde não se tem informação da presença de sítios, totalizando 56 pontos de amostra. Neste modelo, foram selecionados apenas os sítios de manchas de solo escuro, pois, além do fato de ter sido o tipo de sítio mais recorrente

identificado na prospecção arqueológica, este tipo de sítio ocorre em diversas outras áreas da encosta catarinense e pode estar relacionado com um contexto de contato entre grupos Jê e grupos Umbu. Os sítios com mancha de solo escuro ainda são pouco conhecidos pela arqueologia e podem auxiliar na compreensão de diversos fenômenos complexos de contato étnico que ocorrem por volta do ano 1.000 AD na encosta catarinense (FARIAS et al., 2018), portanto um modelo preditivo desta categoria de sítio permite explorar questões sobre este importante período da ocupação humana na região. Todos os valores foram arrolados em uma planilha para a análise estatística no *software* STATGRAPHICS Centurion 16. Foram realizadas três etapas de testes estatísticos de regressão logística, e em cada etapa as variáveis cujo valor-p não alcançavam o grau de 95% de confiança foram removidas. As Tabelas 4, 5 e 6 apresentam o processo dos testes estatísticos, culminando na seleção das variáveis declividade e elevação como sendo as mais explicativas para o padrão de distribuição dos sítios na área de estudo. Os valores-p em destaque são aqueles que foram removidos para o teste estatístico seguinte.

Tabela 4 - Primeiro teste estatístico de Regressão Logística. Fonte: Kozlowski (2018).

Variável	Qui-quadrado	Df	Valor-P
Declividade	13,5502	1	0,0002
Elevação	5,89701	1	0,0152
Radiação Solar	0,47492	1	0,4806
Distância da água	0,226321	1	0,6343

Tabela 5 - Segundo teste estatístico de Regressão Logística. Fonte: Kozlowski (2018).

Variável	Qui-quadrado	Df	Valor-P
Declividade	13,3555	1	0,0003
Elevação	5,73322	1	0,0166
Radiação Solar	0,397177	1	0,5286

Tabela 6 - Terceiro teste estatístico de Regressão Logística. Fonte: Kozlowski (2018).

Variável	Qui-quadrado	Df	Valor-P
Declividade	15,0121	1	0,0001
Elevação	6,75863	1	0,0093

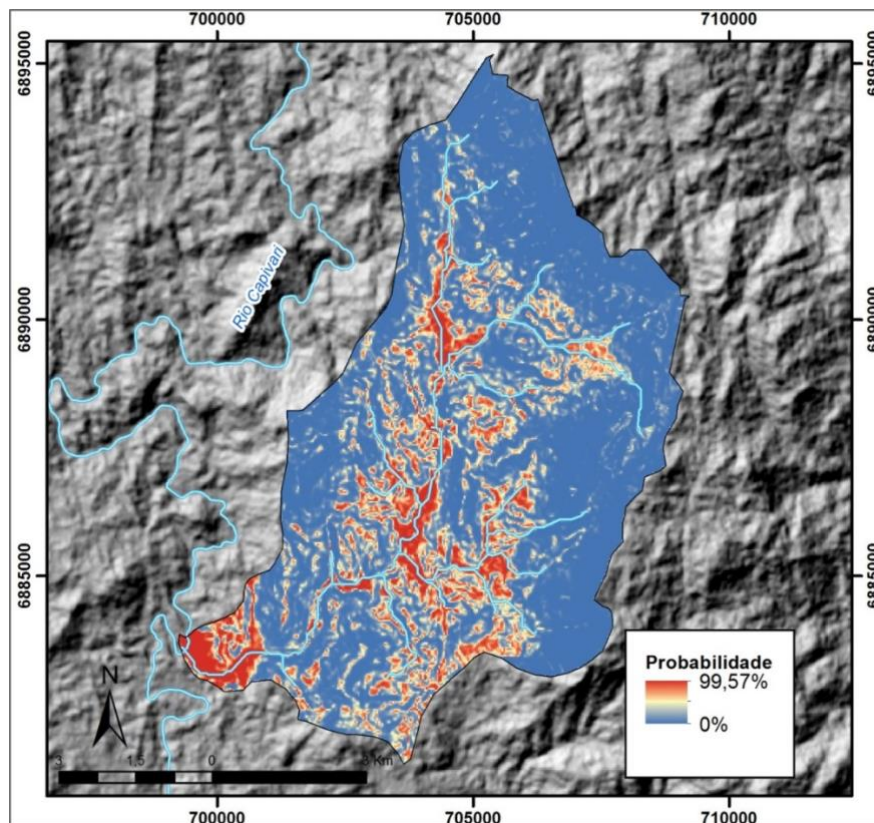


Figura 3 - Modelo preditivo de abordagem dedutiva.

Para apresentar a probabilidade da ocorrência de sítios com manchas, as análises estatísticas resultam em uma equação, a saber:

$$p = \frac{e^{5,7428 - 0,310512 \times Declividade - 0,0101505 \times Elevação}}{1 + e^{5,7428 - 0,310512 \times Declividade - 0,0101505 \times Elevação}}$$

Esta equação é inserida no SIG através da ferramenta *Raster Calculator*, resultando no mapa da Figura 3, onde foi possível identificar áreas de 0% até 99,57% de probabilidade de ocorrência de sítios arqueológicos.

Testando os modelos

Como citado anteriormente, a premissa das modelagens preditivas nesta pesquisa foi a de explorar a área de estudo de acordo com a probabilidade de ocorrência de sítios arqueológicos em diferentes locais e/ou compartimentos do vale, ainda que os modelos

produzidos apresentem grandes limitações em sua capacidade preditiva. A capacidade preditiva, também conhecida como performance, é definida como o grau em que o modelo é capaz de prever de forma correta a ausência e presença de vestígios arqueológicos (VERHAGEN, 2007, p. 116).

Uma forma de avaliar modelos preditivos é através do ganho estatístico de Kvamme (1988). Este método, embora possa ser bastante criticado, permite levantar a discussão comparativa entre modelos a partir da avaliação de dois aspectos objetivos e quantificáveis: precisão e acurácia. A precisão de um modelo preditivo está relacionada com a porcentagem da área total de estudo que é ocupada pelas diferentes zonas de probabilidade. Portanto, um modelo em que as zonas de alta probabilidade de ocorrência de sítios representem apenas 10% da área total é mais preciso do que um modelo onde as áreas de alta probabilidade ocupam 90% da área total. A acurácia, por sua vez, é definida como sendo a quantidade de sítios conhecidos que estão inseridos dentro das zonas determinadas como sendo de alta probabilidade. Precisão e acurácia não são dependentes, nem todo modelo é de alta precisão e alta acurácia. Podem existir modelos altamente precisos, porém pouco acurados e vice-versa. Idealmente, um modelo tenta ser o mais preciso e o mais acurado possível, ou seja, definindo porções restritas da área de estudo como sendo de alta probabilidade e que todos os sítios já conhecidos estejam inseridos dentro delas (BANNING, 2002; VERHAGEN, 2007).

O cálculo de ganho é uma forma de relacionar a precisão e a acurácia e que apresenta a performance de um modelo frente à situação real conhecida. Quanto mais próximo de 1, melhor o desempenho da capacidade preditiva do modelo; quanto mais próximo de 0, pior o desempenho e menor a utilidade (KVAMME, 1988). É possível ter valores negativos no teste de ganho, que, para modelos binários, isto é, que avaliam apenas a presença ou ausência de sítios, é o pior resultado possível. Entretanto, em modelos de diferentes categorias de probabilidade, como é o caso dos modelos deste estudo, as áreas que são definidas como sendo de baixa probabilidade de ocorrência irão apresentar valores negativos de ganho (VERHAGEN, 2007, p. 121). A fórmula de ganho estatístico está apresentada abaixo:

$$\text{Ganho} = 1 - \frac{\text{Precisão}}{\text{Acurácia}}$$

Foram realizados dois diferentes testes de ganho: o primeiro deles foi elaborado para o modelo de abordagem indutiva da área de estudo (Tabela 7) e o segundo teste foi elaborado para o modelo de abordagem dedutiva da área de estudo (Tabela 8). O modelo de abordagem dedutiva, por apresentar um resultado quantitativo de probabilidade, demandou o estabelecimento de três classes de probabilidade, com base nos valores quantitativos: baixa

(0% a 33% de probabilidade), média (33% a 66% de probabilidade) e alta (66% a 99% de probabilidade).

Tabela 7 - Teste de ganho para o modelo de abordagem indutiva – Bacia do Rio Capivaras. Fonte: Kozlowski (2018).

Probabilidade	Área (km²)	Sítios	Precisão	Acurácia	Ganho
Alta	9,12	23	13,27	88,46	0,85
Média	7,78	3	11,31	11,54	0,02
Baixa	51,85	0	75,42	0	-
Total	68,75	26	100	100	0

Tabela 8 - Teste de ganho para o modelo de abordagem dedutiva – Bacia do Rio Capivaras. Fonte: Kozlowski (2018).

Probabilidade	Área (km²)	Sítios	Precisão	Acurácia	Ganho
Alta	23,35	24	33,97	92,31	0,63
Média	25,44	2	37,00	7,69	-3,81
Baixa	19,96	0	29,03	0	-
Total	68,75	26	100	100	0

Ao comparar os testes realizados para os dois modelos, é possível visualizar que o modelo preditivo de abordagem indutiva apresenta um valor de ganho muito maior (0,85) para as áreas de alta probabilidade do que nestas mesmas áreas no modelo de abordagem dedutiva (0,63). O modelo indutivo é bastante preciso e bastante acurado nas áreas de alta probabilidade, enquanto que o modelo dedutivo, ainda que seja bastante acurado, é bem menos preciso, pois cada classe ocupa aproximadamente 1/3 da área total. Em ambos os modelos, as áreas de média e baixa probabilidade apresentam valores pequenos de ganho, como já era esperado, pois as áreas mais interessantes para o modelo serão logicamente aquelas em que há maior probabilidade de ocorrência de sítios.

Modelagem como ferramenta de interpretação

Ao comparar ambos os modelos, ainda que atribuam diferentes valores para as variáveis, é possível perceber resultados visualmente semelhantes, como por exemplo a maior probabilidade de ocorrência de sítios em áreas de fundo de vale, áreas aplainadas de antigos

terraços fluviais. Entretanto, há também discrepâncias nos resultados. De acordo com os testes de ganho, o modelo de abordagem dedutiva apresenta um resultado de menor precisão, pois, ao avaliar o tamanho das áreas de cada uma das classes do modelo, vemos que cada uma representa um terço do total, enquanto que o modelo indutivo apresenta um resultado bem mais preciso. Contudo, é importante ressaltar que o modelo de abordagem dedutiva foi elaborado antes das atividades de prospecção, e baseia-se principalmente em pesquisas anteriores nas áreas de entorno (FARIAS *et al.*, 2018). O modelo de abordagem indutiva, por sua vez, foi produzido após a etapa de prospecção e é baseado em apenas um tipo de sítio, aqueles com mancha de solo escuro, resultando, portanto, na definição de áreas que apresentam características mais específicas do que as áreas definidas pela modelagem dedutiva.

Uma grande vantagem do método de abordagem indutiva é a possibilidade de que a equação gerada para o modelo possa ser extrapolada para áreas maiores do que a área original de pesquisa. Neste sentido, o modelo da bacia do Rio Capivaras foi expandido para toda paisagem da encosta da serra na bacia do Rio Tubarão (Figura 4), permitindo visualizar outras áreas de possível ocorrência de sítios com manchas de solo escuro na região.

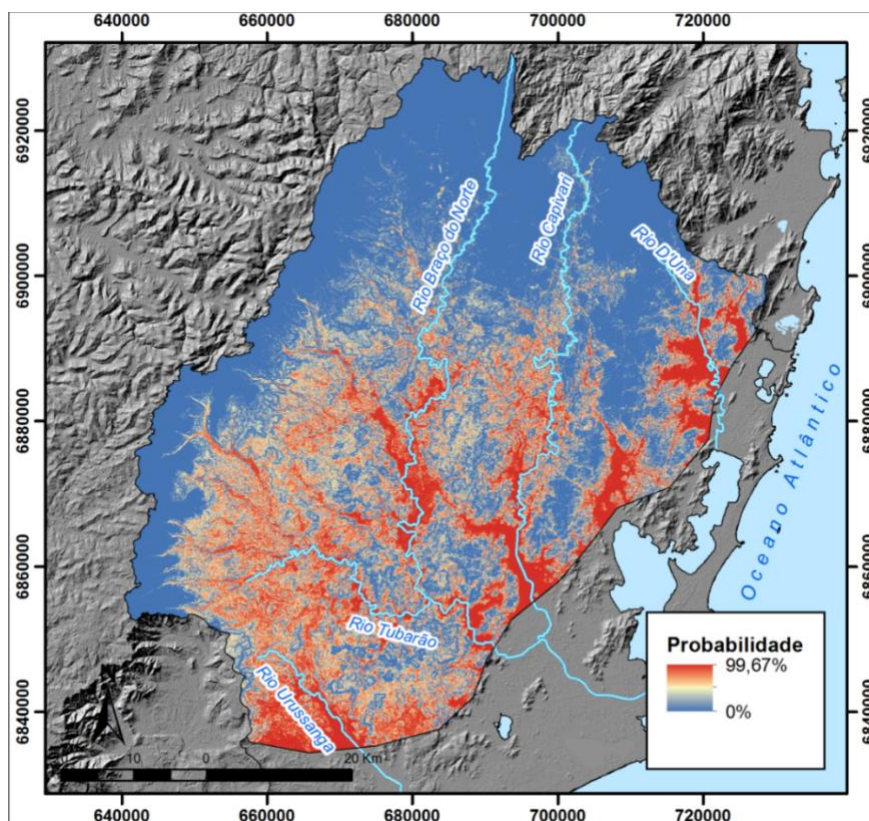


Figura 4 - Modelo preditivo de sítios com mancha de solo escuro na encosta catarinense.

Também foram realizados testes de ganho de Kvamme (1988) para este modelo expandido para toda a região da encosta, no trecho da bacia do rio Tubarão (Tabela 9). Aqui foram também incluídos todos os sítios com mancha de solo escuro conhecidos, registrados pelo GRUPEP-Unisul. Os testes de Kvamme resultaram em um coeficiente de ganho de 0,66 para as áreas de alta probabilidade, com 39 sítios já conhecidos localizados nestas áreas, demonstrando grande acurácia e precisão do modelo preditivo para a bacia inteira (Figura 5).

Tabela 9 - Teste de ganho para o modelo de abordagem indutiva – Bacia do Rio Tubarão. Fonte: Kozlowski (2018)

Probabilidade	Area (km ²)	Sítios	Precisão	Acurácia	Ganho
Alta	1252,25	39	25,75	75,00	0,66
Média	692,14	9	14,23	17,31	0,18
Baixa	2918,35	4	60,01	7,69	-6,80
Total	4862,74	52	100	100	0

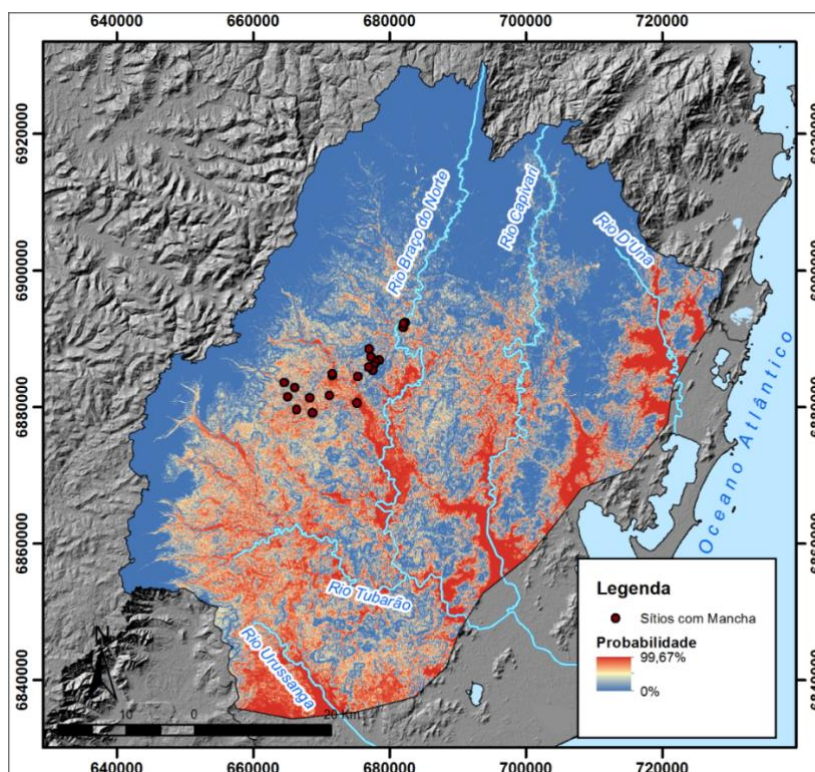


Figura 5 - Sítios com manchas de solo escuro conhecidos fora da área de estudo. Fonte: Kozlowski (2018).

O modelo privilegia as áreas de menor declividade e localizadas nos fundos de vale, próximas aos rios principais da bacia, áreas com maior probabilidade de ocorrência de sítios com manchas de solo escuro. As discussões recentes sobre este tipo de sítio na região têm relacionado estes registros com possíveis processos de etnogênese ligados ao contato entre

grupos Jê e grupos de tecnologia Umbu da encosta (FARIAS *et al.*, 2018; KOZLOWSKI, 2018). Ao associar os sítios caracterizados pela presença de material da Tradição Umbu identificados pelo GRUPEP-Unisul com o modelo proposto para toda a bacia do Rio Tubarão (Figura 6), vemos que a maior parte dos sítios (51,45%) está localizada nas áreas de alta probabilidade do modelo (Gráfico 1).

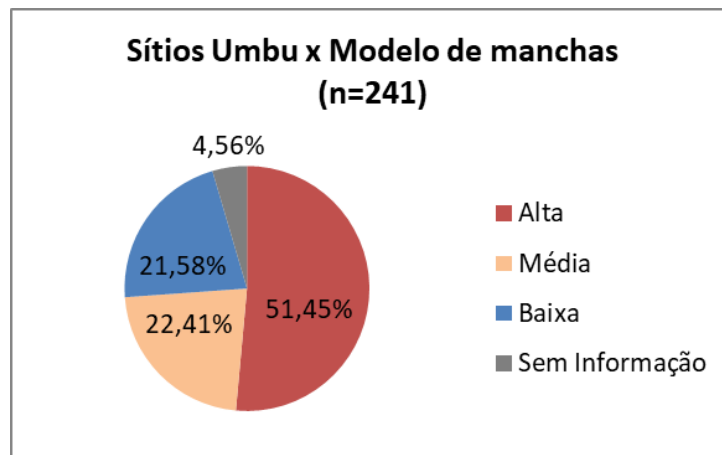


Gráfico 1 - Localização dos sítios Umbu no modelo de sítios com manchas da bacia do Rio Tubarão

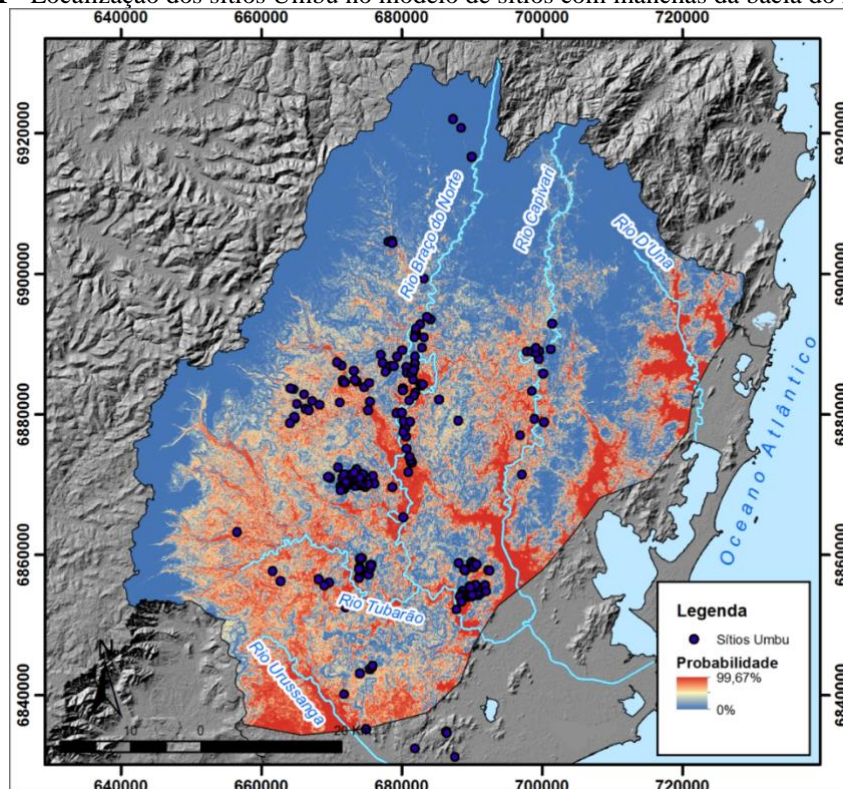


Figura 6 - Sobreposição de sítios Umbu no modelo preditivo. Fonte: Kozlowski (2018).

As áreas terraceadas de fundo de vale correspondem também àquelas preferidas dos sítios da Tradição Umbu. Portanto, conclui-se que há um compartilhamento de características ambientais de localização entre os sítios da Tradição Umbu e os sítios de manchas de solo

escuro. Contudo, a presença de 48,55% dos sítios Umbu em áreas de média e baixa probabilidade indicam que o padrão de distribuição Umbu é mais amplo, extrapolando os parâmetros utilizados neste estudo, baseado nos sítios com manchas de solo escuro.

Considerações finais

Os resultados da modelagem preditiva aplicada na bacia do rio Capivaras trouxe subsídios para a interpretação da distribuição espacial de vestígios arqueológicos, contribuindo também para um desenho mais eficiente de estratégias de prospecção. Entretanto, é necessário assumir uma postura crítica e reconhecer as falhas e limites dos modelos preditivos. A modelagem preditiva não é uma panaceia para acabar com todos os possíveis problemas arqueológicos de prospecção, mas uma ferramenta auxiliar ao longo do processo. Segundo Warren e Asch (2000), a modelagem preditiva é uma ferramenta de projeção de padrões ou relações conhecidas em tempos e lugares desconhecidos, sendo importante destacar a palavra *projeção*, pois, na realidade, os modelos trabalham com correlações apriorísticas cujo sentido para as populações que produziram os vestígios arqueológicos estudados, em última análise, não é conhecido; trata-se, portanto, de aproximações analíticas (EBERT, 2000).

Neste artigo, exploramos os dois métodos de modelagem preditiva mais recorrentes na literatura, avaliando as suas respostas de acordo com testes estatísticos de ganho de Kvamme (1988) e também visualmente, com base em experiências de campo. Analisando visualmente os resultados, podemos interpretar que ambos os métodos propuseram modelos de distribuição espacial muito semelhantes. Contudo, ao analisar estatisticamente através dos coeficientes de ganho, precisão e acurácia, vemos que o modelo de abordagem indutiva apresenta melhor desempenho – lembrando que a análise de ganho pode ser criticada por testar apenas a consistência interna dos modelos, criando muitas vezes uma lógica circular (EBERT, 2000; RIRIS, 2010).

Uma grande vantagem da modelagem preditiva indutiva é a possibilidade de expansão da equação do modelo para uma área mais abrangente, projetando as correlações existentes em uma área piloto e evidenciando outras áreas que podem ser estudadas. Neste sentido, identificamos outras regiões de possível ocorrência de sítios com manchas de solo escuro, que têm sido objeto de interesse de pesquisas regionais para entendimento de processos de ocupação da paisagem da encosta catarinense, constatando também que os sítios com

manchas de solo escuro dispersos por toda a bacia do rio Tubarão apresentam distribuição espacial semelhante à que foi encontrada na bacia do Rio Capivaras.

Este é apenas um estudo exploratório, havendo ainda a necessidade de realizar muitos trabalhos de modelagem preditiva na arqueologia da região e além, pois esta é uma estratégia de pesquisa que ainda se desenvolve de maneira tímida no Brasil. Neste artigo, buscamos dar alguns passos neste campo, ressaltando sempre a importância de analisar criticamente toda e qualquer modelagem, entendendo que um modelo não deve ser algo estático, mas que deve ser constantemente realimentado e reorganizado.

Agradecimentos

Agradecemos ao CNPq e à FAPESP (projeto 2012/51328-3) pelo financiamento da pesquisa, ao GRUPEP-Unisul pelo apoio constante no trabalho de campo, nas análises e na disponibilidade dos dados, ao Museu de Arqueologia e Etnologia da Universidade de São Paulo, a Fabiana Merencio pela ajuda com os modelos e a Bruno Labrador pela ajuda na prospecção e a todos que estiveram envolvidos direta ou indiretamente nas atividades desenvolvidas ao longo deste projeto.

Referências

ARAUJO, Astolfo Gomes de Mello. *Teoria e Método em Arqueologia: Um estudo de caso no alto do Paranapanema, Estado de São Paulo*. Universidade de São Paulo, 2001.

BANNING, E. B. *The Archaeologist's Laboratory - The Analysis of Archaeological Data*. Kluwer Academic Publishers, New York, 2002.

EBERT, James I. The State of the Art in “Inductive” Predictive Modelling: Seven Big Mistakes (and Lots of Smaller Ones). In *Practical Applications of GIS for Archaeologists*, organizado por Konnie L Wescott e R Joe Brandon, pp. 160. Taylor & Francis, London, 2000.

EBLE, Alrpino; REIS, Maria José. Patrimônio Pré-Histórico. In *Relatório Parque Estadual da Serra do Tabuleiro. Aspectos culturais e sociais*, organizado por FATMA/UFSC, 1976. p. 5-45. Florianópolis.

FARIAS, Deisi Scunderlick Eloy; DEMATHÉ, Alexandro; GUIMARÃES, Geovan Martins; MACHADO, Guilherme B.; KNEIP, Andreas; DEBLASIS, Paulo. Aspectos da ocupação da encosta da serra sul catarinense. *Cadernos do LEPAARQ*, 2018. (No prelo).

FARIAS, Deisi Scunderlick Eloy; KNEIP, Andreas. *Panorama Arqueológico de Santa Catarina*. Tubarão: Editora UNISUL, 2010.

FONSECA JÚNIOR, João Aires de 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* 8(3):675–690, 2013.

HODDER, Ian; ORTON, Clive. Spatial analysis in archaeology. In *Spatial analysis in*

archaeology, organizado por Ian Hodder e Clive Orton, pp. 1–16. 1. ed. Cambridge University Press, Cambridge, 1976.

JUDGE, W. Jame; SEBASTIAN, Lynne (Orgs.). *Quantifying the Present and Predicting the Past - Theory, Method, and Application of Archaeological Predictive Modeling*. 1. ed. U.S. Department of the Interior, Bureau of Land Management Service Center, Denver, 1988.

KASHIMOTO, Emília M. O uso de variáveis ambientais na detecção e resgate de bens pré-históricos em áreas arqueologicamente pouco conhecidas. In *Atas do Simpósio sobre Política Nacional do Meio Ambiente e Patrimônio Cultural*, organizado por Solange Bezerra Caldarelli, p. 120–126. Instituto Goiano de Pré-História e Antropologia, Goiânia, 1997.

KIPNIS, Renato. O uso de modelos preditivos para diagnosticar recursos arqueológicos em áreas a serem afetadas por empreendimentos de impacto ambiental. In *Atas do Simpósio sobre Política Nacional do Meio Ambiente e Patrimônio Cultural*, organizado por Solange Bezerra Caldarelli, p. 39–50, 1997. Instituto Goiano de Pré-História e Antropologia, Goiânia.

KNEIP, Andreas. O povo da lagoa: uso do SIG para modelamento e simulação na área arqueológica do Camacho. Universidade de São Paulo, 2004.

KOHLER, Timothy A.; PARKER, Sandra C. Predictive Models for Archaeological Resource Location. *Advances in Archaeological Method and Theory* 9:397–452, 1986.

KOZLOWSKI, Henrique de Sena. Modelagem Preditiva e Ocupação Jê Meridional na Encosta Catarinense. Universidade de São Paulo, 2018.

KVAMME, Kenneth L. Development and testing of quantitative models. In *Quantifying the present and predicting the past. Theory, method and application of archaeological predictive modelling*, organizado por W. James Judge e Lynne Sebastian, p. 325–428, 1988. U.S. Department of the Interior, Bureau of Land Management Service Center, Denver.

KVAMME, Kenneth L. The fundamental principles and practice of predictive archaeological modelling. In *Mathematics and information science in archaeology: A flexible framework*, organizado por A. Voorrips, p. 257–295, 1990. Holos Verlag, Bonn.

MERENCIO, Fabiana T. Acampamentos Xetás no Noroeste do Paraná: o uso da modelagem preditiva para indicação de áreas potenciais para identificação de sítios arqueológicos no território Xetá. 2013.

PLOG, Fred; HILL, J. N. Explaining Variability in the Distribution of Sites. In *The Distribution of Prehistoric Population Aggregates*, organizado por G. J. Gumerman, pp. 7–36. Prescott College Press, Prescott, 1971.

RIRIS, Philip. Predictive modelling in Misiones Province , Argentina using GIS : a case study of the Taquara / Itararé tradition. University of Exeter. 2010.

THOMAZ, Rosângela C. C. O uso de SIG na predição da localização de sítios arqueológicos: um estudo de caso na bacia do Paraná superior. Universidade de São Paulo, 2002.

TILLEY, Christopher. *A Phenomenology of Landscape - Places, Paths and Monuments*. Berg Publishers, Providence, 1994.

TSUCHIYA, Ítalo; TACHIBANA, Vilma Mayumi; IMAI, Nilton Nobuhiro; THOMAZ, Rosângela C. C. Utilização de regressão logística para avaliação de prováveis áreas de ocupações pretérias. *Anais do Simpósio Brasileiro de Geomática*:565–573, 2002.

VERHAGEN, Philip. *Case Studies in Archaeological Predictive Modelling*. Leiden

University Press, Leiden, 2007.

WARREN, Robert E.; ASCH, David L. A Predictive Model of Archaeological Site Location in the Eastern Prairie Peninsula. In *Practical Applications of GIS for Archaeologists*, organizado por Konnie L Wescott e R Joe Brandon, p. 160. Taylor & Francis, London, 2000.

WESCOTT, Konnie L.; BRANDON, R. Joe. *Practical Applications of GIS for Archaeologists*. Organizado por Konnie L Wescott e R Joe Brandon. Taylor & Francis, London, 2000.

WHEATLEY, David; GILLINGS, Mark. *Spatial Technology and Archaeology. The Archaeological Applications of GIS*. Taylor & Francis, Abingdon, UK, 2002.

WILLEY, Gordon R. *Prehistoric Settlement Patterns in the Virú Valley, Peru. Bulletin of the Bureau of American Ethnology*. Vol. 155. Government Printing Office, Washington, 1953.

Nota

i EarthExplorer – USGS: <https://earthexplorer.usgs.gov/>