

**ANALIZA ŚRODOWISKOWYCH UWARUNKOWAŃ
LOKALIZACJI OSADNICTWA ZE STARSZEJ I ŚRODKOWEJ
EPOKI KAMIENIA NA POJEZIERZU LUBUSKIM.
ZASTOSOWANIE DRZEW
KLASYFIKACYJNO-REGRESYJNYCH**

ANALYSIS OF ENVIRONMENTAL VARIABLES
OF PALAEOLITHIC AND MESOLITHIC SETTLEMENT LOCATION
IN POJEZIERZE LUBUSKIE.
APPLICATION OF CLASSIFICATION AND REGRESSION TREES

Iwona Sobkowiak-Tabaka

Instytut Archeologii i Etnologii PAN, Ośrodek Studiów Pradziejowych i Średniowiecznych
ul. Rubież 46, 61-612 Poznań, Poland
iwona.sobkowiak@iaepan.poznan.pl

Jarosław Jasiewicz

Instytut Geoekologii i Geoinformacji, Uniwersytet im. Adama Mickiewicza w Poznaniu
ul. Dziegiełowa 27, 61-680 Poznań, Poland
jarekj@amu.edu.pl

ABSTRACT. The article presents results of the study of environmental variables influencing location of Palaeolithic and Mesolithic settlement in Pojezierze Lubuskie. A new method of predicting archaeological sites location has been suggested, based on classification and regression trees (CART), including nine natural variables. A detailed model has been designed illustrating dependency between these natural variables and intensity of prehistoric activity. The study, results have been also presented on the map, which systematises and presents, in a coherent manner, a model of human presence in Palaeolithic and Mesolithic periods.

KEY WORDS: Pojezierze Lubuskie, Stone Age, environmental variables of settlement, classification and regression trees (CART).

Rozpoznanie uwarunkowań lokalizacji osadnictwa jest szczególnie istotne w odniesieniu do rejonów trudno dostępnych pod względem powierzchniowej prospekcji archeologicznej (łąki, tereny zalesione, torfowiska, nieużytki), w jakie obfi-

tuje młodoglacjalna strefa Niżu Polskiego. Z powodu braku możliwości tradycyjnego rozpoznania powierzchniowego zbadanie struktury osadnictwa późnopaleolitycznego jest poza zasięgiem innych metod nieinwazyjnych, takich jak metoda elektrooporowa, magnetyczna czy elektromagnetyczna, nastawionych na rejestrację większych konstrukcji i obiektów o stosunkowo dużej kubaturze i głębokości (Misiewicz 2006).

Należy podkreślić, że informacje przechowywane w archeologicznych zbiorach danych nie są reprezentatywne. Obszary objęte programami badawczymi zawierają znacznie więcej stanowisk niż te, na których przeprowadzono jedynie podstawowy rekonesans. W tej sytuacji analiza przestrzenna z wykorzystaniem narzędzi modelowania statystycznego staje się jedyną alternatywą dla inwazyjnych sposobów lokalizacji potencjalnych stanowisk. Działanie to pozwala przede wszystkim na opracowanie przestrzennego modelu aktywności społeczności pradziejowych w krajobrazie. Może być również wykorzystywane przez instytucje do selekcji obszarów potencjalnie istotnych dla dziedzictwa kulturowego, na których, z powodów wymienionych powyżej, nie zawsze występują ślady osadnictwa w postaci zgromadzonych artefaktów.

Obszar Pojezierza Lubuskiego, a zwłaszcza jego centralna część obejmująca Pojezierze Łagowskie, należy do najslabiej zbadanych regionów zachodniej Polski, zarówno pod względem osadnictwa późnoglacjalnego, jak i rekonstrukcji ówczesnie panujących warunków środowiska przyrodniczego (Okuniewska-Nowaczyk 2005; Okuniewska-Nowaczyk, Sobkowiak-Tabaka 2014). Rozpoznanie dynamiki zasiedlenia tego terenu przez społeczności późnoglacjalne zmienia się dopiero w ostatnich latach – dzięki szerokopłaszczyznowym badaniom ratowniczym, prowadzonym w związku z rozbudową sieci dróg szybkiego ruchu (Kabaciński, Sobkowiak-Tabaka 2010, 2011, 2013) oraz programom naukowym. Jednym z takich programów jest prowadzony od 2011 r. projekt dotyczący osadnictwa późnoglacjalnego w rejonie Lubrzy, gdzie zarejestrowano koncentrację późnopaleolitycznych i mezolitycznych stanowisk archeologicznych skupionych wokół kopalnego zbiornika wodnego (Sobkowiak-Tabaka, Okuniewska-Nowaczyk, Ratajczak-Szczerba [w druku]). Biorąc pod uwagę szersze studia nad osadnictwem późnoglacjalnym w tej części Polski, dotychczasowy stopień rozpoznania Pojezierza Lubuskiego pod względem archeologicznym oraz niewielką dostępność tego terenu ograniczającą wykonanie badań powierzchniowych, przeprowadzono analizy w zakresie potencjalnych zależności pomiędzy wybranymi cechami środowiskowymi a lokalizacją stanowisk archeologicznych. Podstawowe cele badań były następujące: a) rozpoznanie rozkładu potencjalnej aktywności społeczności starszej i środkowej epoki kamienia w odniesieniu do cech krajobrazu; b) projekcja wyników na cały badany obszar oraz c) wyjaśnienie związków pomiędzy rozmieszczeniem stanowisk a cechami terenu: morfometrycznymi, hydrologicznymi i geologicznymi. W dalszej kolejności planowano wy-

znaczenie i scharakteryzowanie, w skali regionalnej, preferowanych obszarów dla osadnictwa ze starszej i środkowej epoki kamienia¹.

CHARAKTERYSTYKA FIZJOGRAFICZNA POJEZIERZA LUBUSKIEGO

Pojezierze Lubuskie od północy i południa ograniczone jest dużymi formami pradolinowymi o równoleżnikowym przebiegu, tj. Pradolina Toruńsko-Eberswaldzką i Warszawsko-Berlińską, od zachodu południkowym odcinkiem doliny Odry, a od wschodu przez dolny bieg rzeki Obry. Rzeźba tego obszaru jest efektem procesów zachodzących podczas plejstocenijskich zlodowaceń oraz wpływu struktur osadów trzeciorzędowych. Ostatecznie została ukształtowana przez lądolód skandynawski, szczególnie w trakcie jego recesji z linii postojowej fazy leszczyńskiej oraz postoju na linii fazy poznańskiej i jego późniejszej deglacjacji. Krajobraz ten obfituje w liczne moreny czołowe, porozcinane rynnami glacialnymi i dolinami sandrowymi, w których powstały jeziora rynnowe, a później równiny akumulacji biogenicznej (Chmal 2003). Region ten zajmuje powierzchnię ok. 5,1 tys. km² i dzieli się na cztery mezoregiony: Lubuski Przełom Odry, Pojezierze Łagowskie, Równinę Torzymską i Bruzdę Zbąszyńską (Kondracki 2009, s. 134).

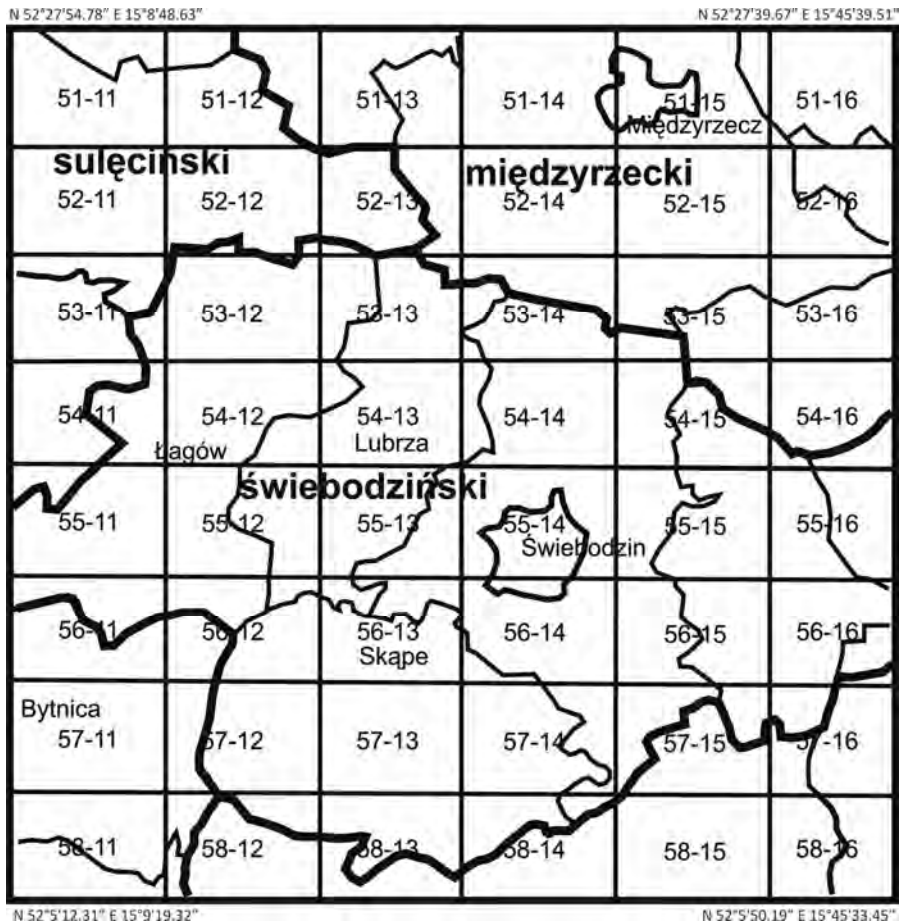
METODA BADAWCZA

Wszelkie dane o stanowiskach archeologicznych z terenu Polski, gromadzone w ramach wieloletniego programu Archeologicznego Zdjęcia Polski (Jaskanis 1992; Konopka 1981), znajdują się na ponad 450 000 Kart Ewidencji Stanowiska Archeologicznego – KESA (Prinke 2002). Wady i zalety aplikacji danych uzyskanych drogą prospekcji powierzchniowej, w szeroko pojętych studiach osadniczych, zostały przedstawione w literaturze przedmiotu (por. np. Jankowska, Rozbiegalski 2012, s. 140–141; Jasiewicz, Hildebrandt-Radke 2009a, 2009b). Warto jednak w tym miejscu zwrócić uwagę na dwa podstawowe problemy wynikające z wykorzystywania danych pochodzących z KESA. Przede wszystkim to bardzo ogólne określenie chronologii stanowiska na podstawie materiałów pozyskanych głównie z badań powierzchniowych. Często ograniczają się one, w szczególności w odniesieniu do artefaktów krzemiennych, do stwierdzenia „epoka kamienia” (por. poniżej). Drugim nie mniej ważnym minusem jest to, że brak stanowisk na danym obszarze AZP nie przekłada się w pełni na stan rzeczywisty. Może to bowiem oznaczać także brak

¹ Badania prowadzono w ramach projektu „Późnoglacialne społeczności łowiecko-zbierackie Pojezierza Lubuskiego. Chronologia, systematyka, podstawy utrzymania”, finansowanego przez Narodowe Centrum Nauki (projekt nr DEC-2011/01/D/HS3/04134).

możliwości ich zewidencjonowania z powodu utrudnionej obserwacji terenowej. Sytuacja ta w znacznej mierze odnosi się do analizowanego w tym artykule Pojezierza Lubuskiego, ponieważ tylko ok. 40% tego obszaru zajmują pola uprawne umożliwiające prospekcję powierzchniową, a ponad 50% pokrywają lasy, łąki i nieużytki.

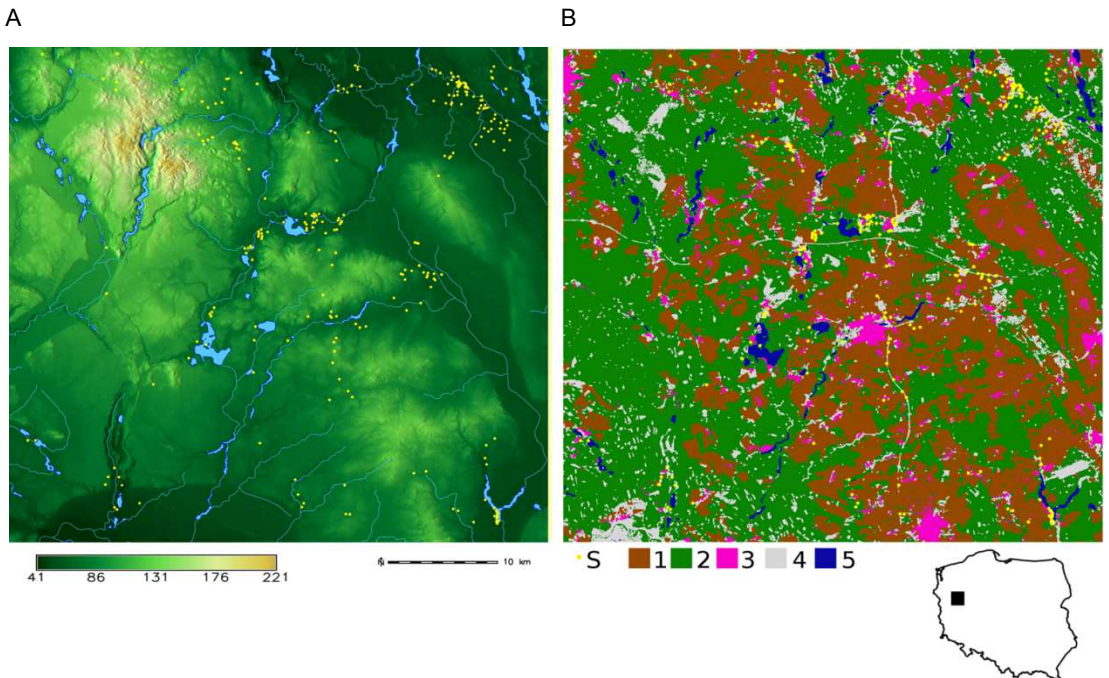
Zaproponowane przez nas rozwiązania analityczne, będące jednym z narzędzi eksploracji danych – drzewa decyzyjno-regresyjne (Classification and Regression Trees – CART) – pozwalają zminimalizować problemy związane z niereprezentatywnością dyspersji stanowisk oraz ich nadreprezentatywnym skupieniem w kilku miejscach, wynikającym często z intensywności prowadzonych badań powierzchniowych.



Ryc. 1. Lokalizacja arkuszy AZP na tle podziału administracyjnego Polski

Fig. 1. Location of Polish Archaeological Record (Polish AZP) sheets on the map of administrative division of Poland

Proces zbierania danych polegał na zestawieniu i cyfryzacji danych z Archeologicznego Zdjęcia Polski obejmującego 48 arkuszy AZP, czyli ok. 1,7 tys. km² (ryc. 1), zdeponowanych w Wojewódzkim Urzędzie Ochrony Zabytków w Zielonej Górze oraz w Delegaturze tegoż Urzędu w Gorzowie Wielkopolskim. Dygitalizację danych, ze względu na obszerność zbiorów, ograniczono jedynie do stanowisk starszej i środkowej epoki kamienia. Uwzględniono również stanowiska młodszej epoki kamienia zawierające znaleziska krzemienne. Ogółem rozpoznano 282 stanowiska, w tym 20 stanowisk zawierających znaleziska paleolityczne i 50 – mezolityczne. Niestety większość z powodu braku charakterystycznych cech technologiczno-typologicznych artefaktów krzemiennych, tj. 184 stanowiska, została sklasyfikowana jedynie jako „epoka kamienia”, bez żadnych możliwości uściślenia chronologii. Ponadto w trakcie zbierania materiału okazało się, że liczba prób (w tym przypadku: stanowisk archeologicznych) nie jest reprezentatywna statystycznie dla różnych typów krajobrazu i uniemożliwia opracowanie modelu regresyjnego, np. według metodyki zaproponowanej przez J. Jasiewicza i I. Hildebrandt-Radke (2009a). W związku z tym zdecydowano się na zastosowanie metod związanych z odkrywa-



Ryc. 2. A – lokalizacja stanowisk na tle mapy fizjograficznej (rzędna i sieć rzeczna) badanego obszaru, B – użytkowanie terenu na obszarze badań: S – lokalizacja stanowisk archeologicznych; 1 – pola orne; 2 – lasy; 3 – obszary zurbanizowane; 4 – łąki; 5 – jeziora

Fig. 2. A – Location of sites on the physiographic map (ordinate and river network) of the studied area, B – usage of the terrain in the study area: S – location of archaeological sites; 1 – arable lands; 2 – forests; 3 – urban areas; 4 – meadows; 5 – lakes

niem wiedzy w bazach danych (Knowledge Discovery in Database), których podstawą jest *data mining* – eksploracja danych (Fayyad, Piatetsky-Shapiro, Smyth 1996). Metody te zamiast złożonych modeli wyszukują jedynie te cechy, które są istotne dla rozwiązania problemu, pomijając nieistotne lub te niedające rozstrzygnięć.

Dane zostały zgromadzone w postaci cyfrowej bazy, w której przechowywano współrzędne stanowiska oraz wszystkie atrybuty uzyskane z bazy AZP. Lokalizację stanowisk na tle rzeźby terenu przedstawiono na rycinie 2A.

Cyfrowy model wysokościowy obszaru ograniczonego współodrędnymi (ryc. 1) został wykonany na podstawie cyfrowego modelu wysokościowego DEDT L2, przekształconego do postaci zmiennoprzecinkowej i układu współrzędnych ETRS92 o rozdzielczości 30×30 m, za pomocą autorskiego programu. Dodatkowo w celu odróżnienia obszarów, w obrębie których przeprowadzono badania AZP (np. pola orne, miejscowości), od obszarów, na których takich badań nie prowadzono na skalę masową (lasy, łąki), wykonano uproszczoną mapę użytkowania terenu (ryc. 2B) przygotowaną przez Annę Manikowską i Tomasza Elantkowskiego w ramach prac magisterskich w Zakładzie Geologii i Paleogeografii Czwartorzędu UAM.

WARSTWY GEOPRZESTRZENNE: DANE ARCHEOLOGICZNE I POCHODNE TERENU

Stanowiska archeologiczne mają postać kilku skupień. Ze względu na to, że trudno było rozróżnić, czy są one efektem naturalnego procesu intensywnego użytkowania w przeszłości, czy też jedynie nierównomiernego rozpoznania, zdecydowano się na zabieg zaproponowany przez Williama, Ulma, Cooka, Langleya i Collarda (2013), polegający na generalizacji stanowisk na siatce o niższej rozdzielczości w taki sposób, że każdemu oczku siatki przypisywano 0 – jeżeli uznano, że przeprowadzono tam badania AZP i nie było tam żadnych stanowisk, lub 1 – jeżeli wystąpiła dowolna liczba stanowisk większa niż 0. Obszarom, co do których uważano, że nie przeprowadzono na nich badań AZP, tj. obszarom leśnym i łąkom, przypisywano wartość pustą (NULL). W ten sposób zminimalizowano wpływ zagęszczenia będących jedynie wynikiem dokładniejszego rozpoznania wybranych fragmentów terenu (np. okolice Wojnowa)².

Następnie wyliczono pochodne terenu. Pierwszy etap ich liczenia odbywał się na podstawie opracowanego modelu o rozdzielczości 30 m, później natomiast pochodne sumowano/grupowano w ramach tej samej siatki (o niskiej rozdzielczości), której użyto do generalizacji danych archeologicznych. Pochodne terenu (lista wg tabeli 1

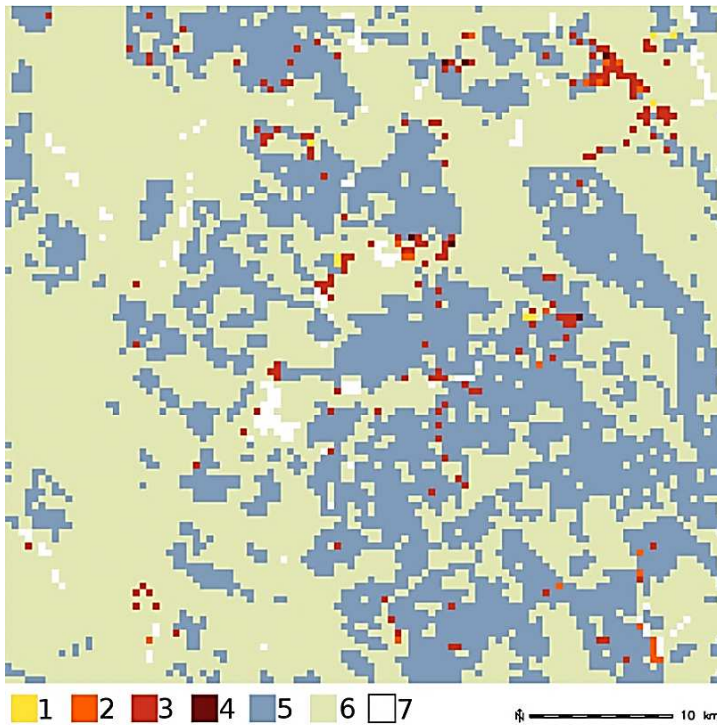
² Wojnowo i jego okolice były od końca lat 60. XX w. objęte szczegółowymi badaniami prowadzonymi pod kierownictwem prof. dr. hab. M. Kobusiewicza (por. np. Bower, Kobusiewicz 2002; Kobusiewicz 1999).

Tabela 1. Pochodne terenu i metody ich wyznaczania
Table 1. Terrain derivatives and methods of their determination

Pochodna	Charakterystyka pochodnej
RELBASE wysokość względem bazy erozyjnej	Względna wysokość powyżej bazy erozyjnej. Powierzchnia bazy została wyznaczona jako powierzchnia łącząca węzły pomiędzy strumieniami 4. i 5. rzędu klasyfikacji A.N. Strahlera (1952) a ich dopływami. Zmienna pozwala rozróżnić wyniesienia i obniżenia niezależnie od ich wysokości względnej.
RANGE wysokość względa, relief	Różnica pomiędzy najniższą i najwyższą wartością rzędnej oryginalnego modelu wysokościowego w danej komórce siatki o niskiej rozdzielczości. Dostarcza informacji na temat otwartości obszaru, szorstkości terenu, może być interpretowana na wiele sposobów – jako koszt przemieszczania, możliwość występowania kryjówek lub wyniesień.
FLATNESS płaskość	Udział (wyrażony wartościami od 0 do 1) obszarów płaskich lub nieznacznie nachylonych (poniżej 4 stopni), liczonych dla oryginalnego modelu w każdej komórce siatki o niskiej rozdzielczości. Zmiana informuje o płaskości obszaru. Im wyższa wartość, tym większa dominacja obszarów płaskich.
TEXTURE tekstura	Suma długości wcięć w każdej komórce siatki. Wcięcia zostały wyliczone z oryginalnego modelu wysokościowego jako formy dolinne, następnie zredukowane do postaci szkieletowej i zsumowane oraz przeliczone do postaci km/km ² (gęstość) dla siatki o niskiej rozdzielczości. Informuje o koszcie przemieszczania oraz dostępności możliwych miejsc ukrycia zwierzyny.
LOWLAND nizinność	Udział obszarów poniżej mediany dla danej komórki siatki. Podaje informację na temat obecności wyniesień w danej komórce. Komórki z dużą wartością są mniej wyniesione (i możliwie mniej widoczne) niż obszary wyniesione.
WETNESS wskaźnik wilgotności potencjalnej	Wskaźnik topograficzny (TCI) liczony bezpośrednio dla siatki o niskiej rozdzielczości. Określa potencjalną wilgotność terenu.
DISTWATER	Liczony jako minimalna hydrologiczna (wzdłuż linii spływu) odległość pomiędzy daną komórką siatki a najbliższym jeziorem lub rzeką. Dla małych jezior (poniżej 1 ha) odległość była liczona tylko dla najbliższej komórki. Jeżeli w odległości 1 km nie występowały żadne obiekty o powierzchni mniejszej niż 1 ha, odległość była liczona do najbliższego dużego obiektu. Zmienna określa koszt dostępności zbiorników wodnych: przede wszystkim w kontekście rybołówstwa.
FORMS	Zmienna kategoryzacyjna. Forma terenu liczona za pomocą metody zaproponowanej przez Jasiewicza i Stepinskiego (2013), wyznaczającej najbardziej popularne formy terenu.
GEOL	Litologia powierzchni; występują trzy klasy: gliny morenowe, piaski fluwioglacjalne i osady organiczne. Mapa opracowana na podstawie <i>Mapy geologicznej Polski</i> w skali 1:500 000 (Marski in. 2006)

oraz ryc. 4) zostały wyliczone za pomocą metod opisanych w tabeli 1. W ten sposób otrzymano bazę danych zawierającą 10 warstw rastrowych: 9 – obejmujących informację przyrodniczą – traktowanych jako zmienne niezależne pełniące funkcję wyjaśniającą (*explanatory variables*) oraz 1 warstwę zawierającą zmienną zależną, przechowującą informację o obecności lub braku stanowisk archeologicznych albo braku obserwacji (*dependend variable, response variable*). Warstwę zawierającą informację o obecności stanowisk epoki kamienia na tle stopnia rozpoznania terenu przedstawiono na ryc. 3.

Zestaw zmiennych wyjaśniających użyty w analizie obejmował te zmienne morfometryczne, co do których przypuszczano, że mogą mieć wpływ na aktywność społeczności epoki kamienia. Niestety dla starszej i środkowej epoki kamienia trudno ustalić jakiegokolwiek zmienne oparte na relacjach społecznych lub ekonomicznych innych niż reinterpretacja zmiennych przyrodniczych, takich jak np. odległość od



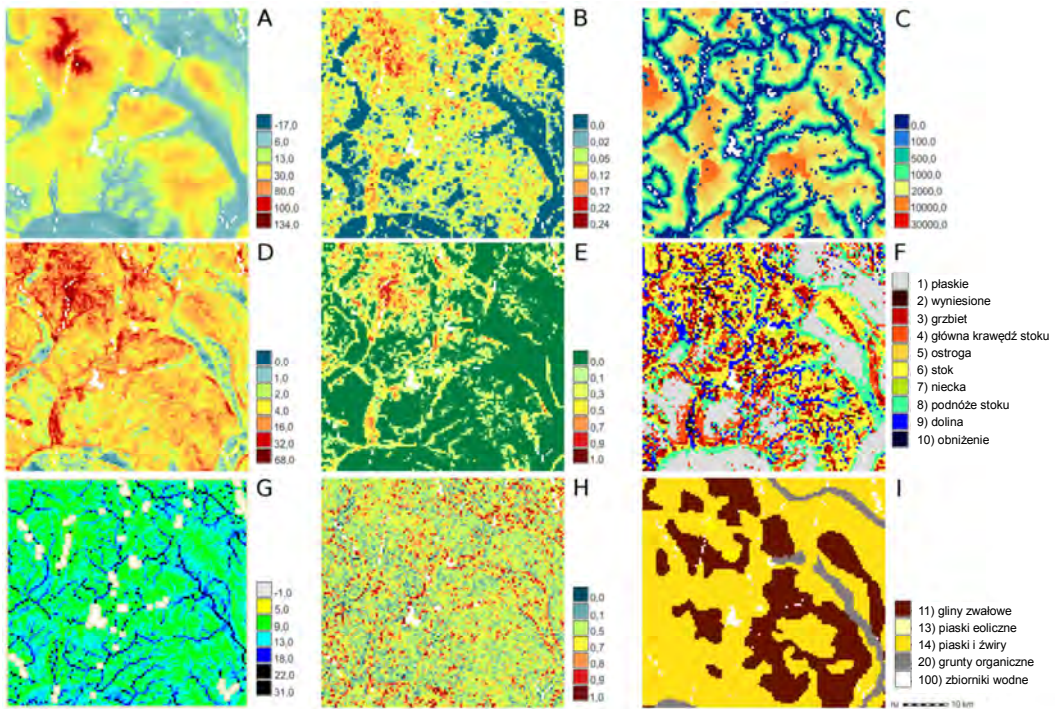
Ryc. 3. Warstwa zawierająca zmienną wyjaśniającą – informacje o obecności stanowisk archeologicznych z epoki kamienia: 1 – stanowiska paleolityczne/z epoki kamienia; 2 – stanowiska mezolityczne/z epoki kamienia; 3 – stanowiska z epoki kamienia; 4 – stanowiska zarówno paleolityczne, jak i mezolityczne/z epoki kamienia; 5 – brak informacji o stanowiskach (potencjalnie istniejące, lecz nieodkryte stanowiska z epoki kamienia); 6 – obszar niezbadany; 7 – jeziora

Fig. 3. A level with explanatory variable – information of a presence of the Stone Age archaeological sites: 1 – Palaeolithic/Stone Age sites; 2 – Mesolithic/Stone Age sites; 3 – unresolved Stone Age sites; 4 – both Palaeolithic and Mesolithic/Stone Age sites; 5 – negative (FALSE) examples (no Stone Age sites detected); 6 – non-inspected area, 7 – lakes

wody (Duke 2003; Vanacker i in. 2001). Brak aktywności rolniczej w tym czasie lub duża jednorodność terenu nie pozwalają na formułowanie w tej materii żadnych wniosków.

DRZEWA DECYZYJNO-REGRESYJNE

Podstawową metodą analizy były drzewa klasyfikacyjno-regresyjne. Jest to jeden z najbardziej popularnych algorytmów uczenia maszynowego (Goldman i in. 1982; Breimani i in. 1984). Są to narzędzia służące do budowy modeli predykcyjnych (predykcja jakościowa lub ilościowa zjawiska) i deskryptywnych (opis i pre-



Ryc. 4. Pochodne terenu (por. tab. 1). A – pozycja topograficzna, B – płaskość, C – odległość od wody, D – różnica wysokości, E – urozmaicenie, F – formy terenu, G – wilgotność, H – nizinność, I – geologia
 Fig. 4. Terrain variables (see also Table 1). A – relbase, B – flatness, C – distwater, D – range, E – texture, F – forms, G – wetness, H – lowlands, I – geol

zentacja wzorców w badanej zbiorowości). Główną zaletą ich stosowania jest to, że nie bazują one na żadnych założeniach statystycznych przyjętych *a priori* (Cook, Goldman 1984).

Ich działanie opiera się na rekursywnym (powtarzającym się) dzieleniu zbioru w taki sposób, aby po każdym podziale „czystość” zmiennej zależnej w poszczególnych podzbiorach (liściach drzewa) była jak największa (Zhang 1998). Zmienna i jej wartość wykorzystywana w podziale jest tak dobierana, aby z grupy kilku zmiennych alternatywnych (*surrogates*) zysk na danym etapie był największy. Szybkie osiągnięcie pełnej czystości (gdy liście drzewa zawierają albo tylko komórki ze stanowiskami archeologicznymi, albo bez stanowisk archeologicznych) w praktyce nie jest możliwe, dlatego drzewa są z reguły kompromisem pomiędzy stopniem ich rozbudowy a osiągniętą ostatecznie skutecznością. W ten sposób liście zawierają wartość proporcji [0–1] pomiędzy liczbą komórek zawierających i niezawierających stanowisk archeologicznych. Proporcja ta traktowana jest jako wartość decyzyjna,

Tabela 2. Schemat konstrukcji macierzy zmieszania

Table 2. Scheme of a confusion matrix generating

Warunek				
	Dokładność $Dokl=(PP+FP)/(PR+P)$	Pozytywny $P=ZP+FN$	Negatywny $N=FP+PN$	
Wynik testu	Pozytywny $P'=PP+FP$	Prawdziwie pozytywny PP	Falszywie pozytywny FP	Pozytywna wartość predykcyjna (precyzja) $Prec=PP/P'$
	Negatywny $N'=FN+PN$	Falszywie negatywny FN	Prawdziwie negatywny PN	Negatywna wartość predykcyjna $nwp=PN/N'$
		Czułość (przywołanie) $Prz=PP/P$	Specyficzność $Spec=PN/N$	Wartość testu statystycznego $F=2*Prec*Prz/(Prec+Prz)$

P – pozytywny; FP – fałszywie pozytywny; PR – prawdziwy; PP – Prawdziwie pozytywny; FN – fałszywie negatywny; PN – prawdziwie negatywny

Tabela 3. Analiza wydajności (modeli) dla różnych wartości progowych (zob. tab. 2 dla objaśnień)

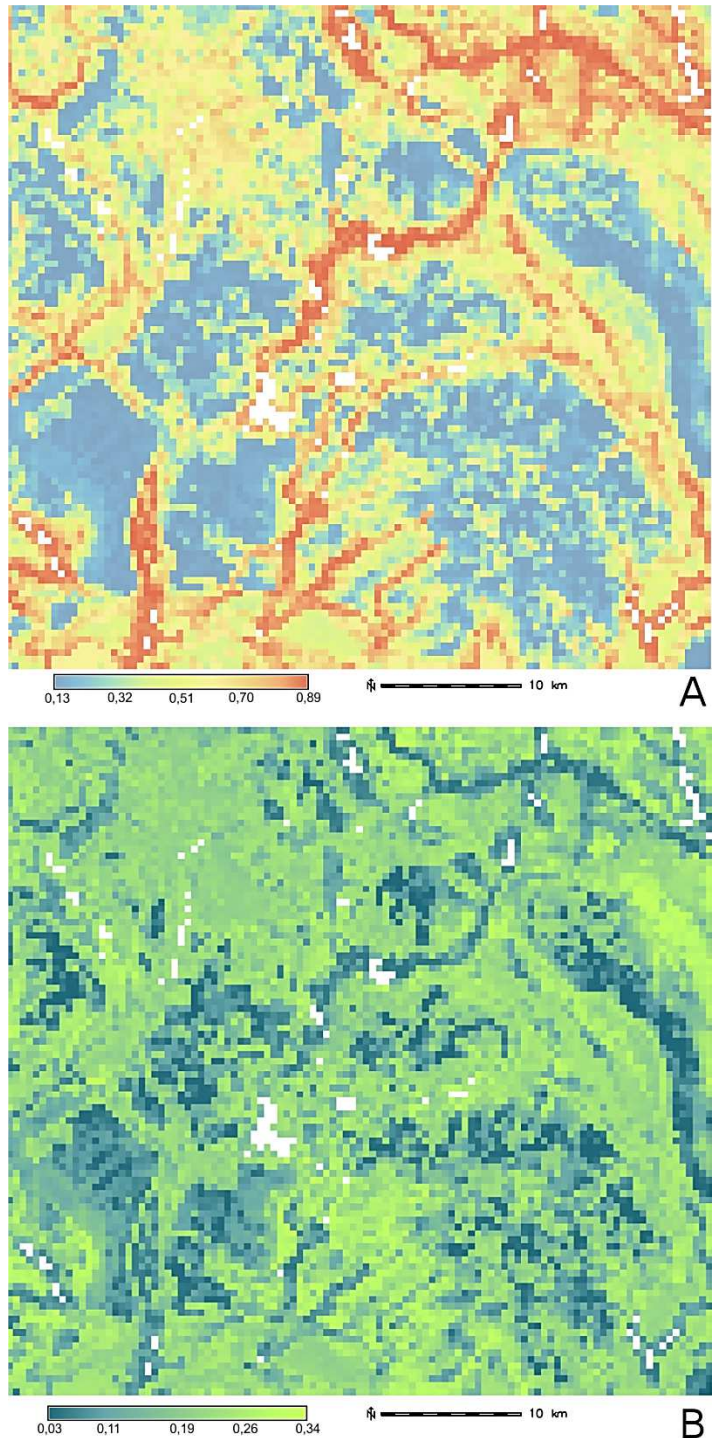
Table 3. Analysis of productivity (of models) based on various limiting values (see Table 2 for explanations)

	Uśredniona wartość decyzyjna+odchylenie standardowe (MEAN+STDDEV)	Uśredniona wartość decyzyjna (MEAN)	Uśredniona wartość decyzyjna – odchylenie standardowe (MEAN-STDDEV)
Zrównoważona dokładność	0,76	0,79	0,69
Pozytywna wartość predykcyjna/precyzja	0,30	0,14	0,09
Negatywna wartość predykcyjna	0,97	0,98	0,99
Czułość/przywołanie/współczynnik odtwarzania prawdziwości	0,60	0,86	0,93
Specyficzność	0,92	0,72	0,47
1 – specyficzność/współczynnik błędu	0,08	0,28	0,53
Wartość testu statystycznego	0,40	0,25	0,16

określająca, czy dana komórka zostanie w konsekwencji zakwalifikowana jako potencjalnie zawierająca lub niezawierająca stanowisk archeologicznych.

Drzewo klasyfikacyjno-regresyjne staje się w ten sposób klasyfikatorem – zestawem reguł, które następnie w trakcie procedury predykcji stosuje się do nowych danych w celu zakwalifikowania każdego nieznanego elementu do jednej z klas występujących w zmiennej zależnej.

Drzewa klasyfikacyjno-regresyjne, podobnie jak inne algorytmy uczenia maszynowego, pracują dobrze, jeżeli istnieje równowaga pomiędzy klasami w zmiennej zależnej. W przypadku kiedy jedna klasa znacząco dominuje nad drugą (rzędu 10×, 100× lub więcej) tego typu algorytmów, pojawia się problem niezbilansowania danych (Estrabrooks i in. 2004; Garcia, Haibo 2009; Japkowicz 2000), który uniemożliwia stosowanie wyżej wymienionych procedur w normalny sposób. Dane



Ryc. 5. Wyniki symulacji pokazujące intensywność osadnictwa w epoce kamienia. A – średnia intensywność dla 200 powtórzeń, B – odchylenie standardowe

Fig. 5. Results of simulation illustrating intensity of the Stone Age settlement. A – an average intensity for 200 repetitions, B – standard deviation

Tabela 4. Liczba stanowisk w poszczególnych klasach i test całkowitej losowości przestrzennej: 1) liczba komórek, 2) liczba przykładów, 3) liczba pozytywnych przykładów, 4) liczba negatywnych przykładów, 5) minimalna wartość przedziału losowości przestrzennej, 6) maksymalna wartość przedziału losowości przestrzennej

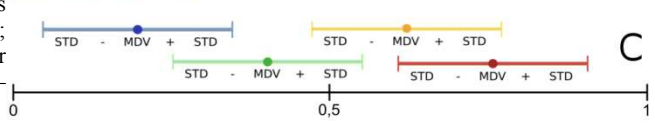
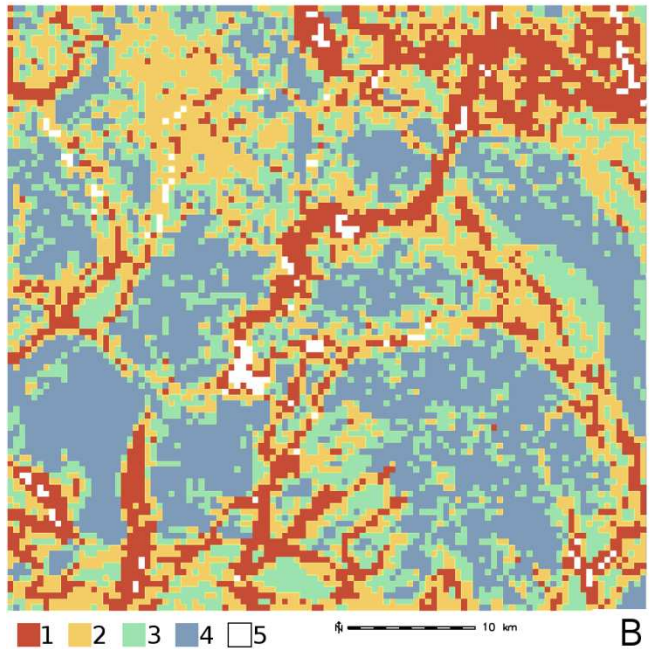
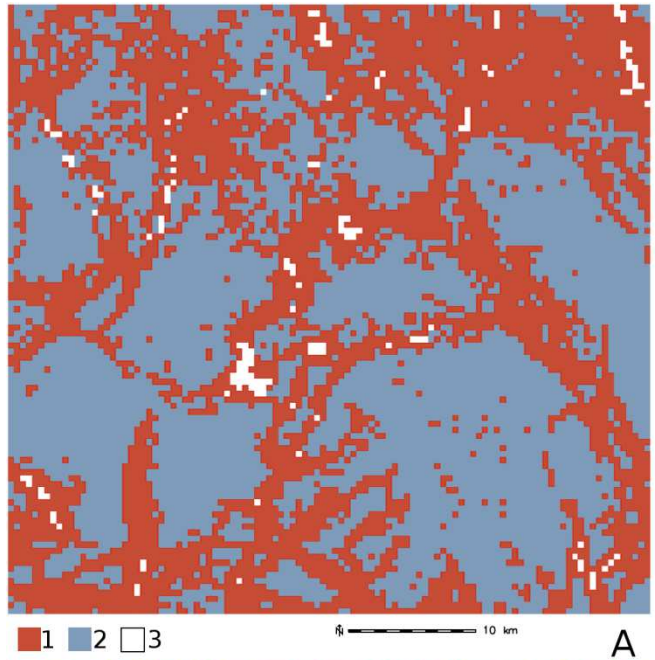
Table 4. Number of sites within particular classes and a complete spatial randomness test: 1) number of grid cells, 2) number of examples, 3) number of positive examples, 4) number of negative examples, 5) minimum spatial randomness value, 6) maximum spatial randomness value

Klasa	1	2	3	4	5	6
suma	10 660	4040	211	3829		
Model prosty (2 klasy)						
negatywne >0,5	5918	2806	29	2777	109	178
pozytywne <0,5	4742	1234	182	1052	49	98
Model rozszerzony (4 klasy)						
zawsze pozytywne (>mean-stddev)	1875	416	126	290	20	35
pozytywne (0,5 – mean+stddev)	2867	818	56	762	29	63
negatywne (mean-stddev – 0,5)	2479	990	14	976	34	63
zawsze negatywne(<mean+stddev)	3439	1816	15	1801	75	115

archeologiczne są typowymi danymi niezbilansowanymi, ponieważ obecność nielicznych stanowisk archeologicznych nie jest równoważona przez znacznie liczniejsze obszary, na których ich nie odkryto pomimo przeprowadzonych badań (por. tab. 4).

W celu rozwiązania problemu zastosowano procedury przywracania balansu danych, polegające na wielokrotnym losowaniu (Estrabrooks i in. 2004) znacznie liczniejszej populacji komórek (3829) niezawierających danych próby losowej równoważnej populacji zawierającej dane (211 komórek). Symulację powtórzono 200 razy, a następnie wyniki predykcji uśredniono oraz wyliczono odchylenie standardowe (ryc. 5A, B). Uśrednione wyniki posłużyły do podjęcia decyzji o zakwalifikowaniu komórki jako potencjalnie zawierającej lub niezawierającej danych archeologicznych.

Test modelu przeprowadzono, stosując standardowe w uczeniu maszynowym macierze zmieszania (*confusion matrices*). Macierze te (tab. 2) pozwalają na ocenę jakości otrzymanego modelu oraz wyliczenie standardowych wskaźników: czułości (*sensitivity*), specyficzności (*specificity*), dokładności (*accuracy*) i indexu F, przedstawionych również w tabeli 3. Wskaźniki te umożliwiają porównywanie modeli otrzymywanych różnymi metodami i z zastosowaniem różnych wartości progowych.



Ryc. 6. Modele prawdopodobieństwa obecności stanowisk archeologicznych. A – model dla dwóch klas: 1 – pozytywnej; 2 – negatywnej; 3 – jeziora; B – model dla czterech klas: 1 – zawsze pozytywne; 2 – pozytywne; 3 – negatywne; 4 – zawsze negatywne; 5 – jeziora

Fig. 6. Models of possible archaeological sites location. A – model for two classes: 1 – positive; 2 – negative; 3 – lakes; B – model for four classes: 1 – always positive; 2 – positive; 3 – negative; 4 – always negative; 5 – lakes

PREDYKCJA MODELI ORAZ TESTY LOSOWOŚCI PRZESTRZENNEJ

Na podstawie uśrednionych wartości decyzyjnych (SWD) oraz ich odchylenia standardowego (STD) zaproponowano dwa modele predykcji obecności stanowisk z epoki kamienia. W pierwszym modelu, zgodnie ze standardami uczenia maszynowego, zastosowano cięcie (*alpha cut*) na poziomie 0,5. Tym samym wyznaczono obszary potencjalnej obecności stanowisk archeologicznych (pozytywne) oraz obszary, na których nie należy się ich spodziewać (negatywne). Wydajność takiego modelu nie jest wysoka, głównie z powodu dużej liczby komórek rozpoznanych fałszywie jako pozytywne (*false positives*). Jest to koszt wynikający z niebilansowania danych. W drugim modelu zastosowano podejście niestandardowe, gdzie w obrębie komórek zakwalifikowanych pozytywnie i negatywnie wyróżniono dodatkowe grupy, progując SWD zmodyfikowane o odchylenie standardowe. Pozwoliło to wyróżnić cztery klasy: *zawsze pozytywne*, *pozytywne*, *negatywne* i *zawsze negatywne*. Klasa *zawsze pozytywne* i *zawsze negatywne* obejmuje te komórki, które w procesie symulacji dla każdej iteracji (powtórzenia) z 95% prawdopodobieństwem były klasyfikowane jako pozytywne i negatywne. Porównanie obu modeli (ryc. 6A, B, tab. 3) wskazuje, że prosty model obejmujący jedynie dwie klasy dla wyników *pozytywnych* prawidłowo przywołuje 182 przykłady pozytywne (*true positives*), ale kosztem 1052 fałszywie pozytywnych. Model oparty na 4 klasach dla klasy *zawsze pozytywne* prawidłowo przywołuje 126 przykładów pozytywnych, ale kosztem 290 przykładów fałszywie pozytywnych. Mimo że przywołuje zaledwie 59% terenów, na których wystąpiły stanowiska z epoki kamienia, jest to obszar obejmujący zaledwie 10% rozpoznanego arealu. Taki wynik wskazuje, że stanowiska występują jedynie na ograniczonym i dobrze zdefiniowanym terytorium.

Aby odpowiedzieć na pytanie, czy wyróżnione klasy potencjalnej obecności stanowisk archeologicznych można utożsamiać z preferowaniem lub unikaniem pewnych typów terenów, przeprowadzono test losowości przestrzennej rozpoznanych stanowisk (Jasiewicz, Hildebrandt-Radke 2009a). Test opiera się na porównaniu liczby komórek, w których stwierdzono obecność stanowisk archeologicznych, z potencjalną liczbą stanowisk danej klasy, jakiej należałoby się spodziewać, gdyby ich rozmieszczenie było przypadkowe. Wyniki testu zostały przedstawione w tabeli 4. Test ten polega na wyliczeniu maksymalnej i minimalnej liczby stanowisk, jakiej można by się spodziewać na obszarze o powierzchni każdej klasy, gdyby ich liczba była losowa. Okazało się że jedynie klasa *zawsze pozytywne* zawiera koncentrację stanowisk większą niż wynika z całkowitej przestrzennej losowości. Klasa *pozytywne* obejmuje obszary obojętne, a klasy *zawsze negatywne* i *negatywne* to tereny o koncentracji stanowisk znacznie poniżej wartości wynikających z testu całkowitej przestrzennej losowości.

Tabela 5. Klasy opisujące związek pomiędzy cechami środowiska a preferencjami osadniczymi: 1) liczba komórek, 2) liczba komórek ze stanowiskami, 3) stosunek pozytywnych do negatywnych w modelu, 4) odchylenie standardowe, 5) średni błąd kwadratowy predykcji, 6) interpretacja

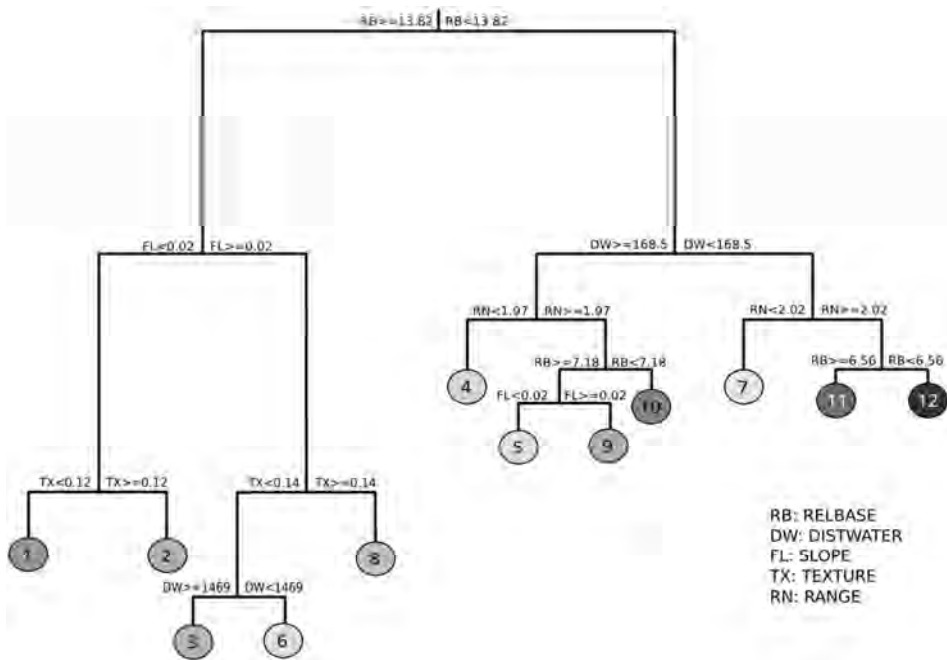
Table 5. Classes describing relationship between environmental features and settlement preferences: 1) number of grid cells, 2) number of grid cells with sites, 3) positive-to-negative ratio, 4) standard deviation, 5) average squared prediction error, 6) interpretation

Klasa	1	2	3	4	5	6
1	2856	9	0,168	0,101	0,004	silnie unikane
2	279	7	0,282	0,220	0,0039	unikane
3	1224	8	0,344	0,206	0,0088	unikane
4	582	4	0,454	0,220	0,0042	obojętne
5	727	14	0,464	0,261	0,0133	obojętne
6	1157	19	0,513	0,237	0,0101	obojętne
7	232	6	0,594	0,244	0,0044	obojętne
8	969	24	0,610	0,205	0,0065	obojętne/preferowane
9	281	10	0,673	0,227	0,0107	preferowane
10	854	27	0,693	0,221	0,0085	preferowane
11	548	11	0,704	0,209	0,021	preferowane
12	948	72	0,839	0,110	0,0044	silnie preferowane
Suma	10657	211				

ANALIZA ZWIĄZKU POMIĘDZY LOKALIZACJĄ STANOWISK ARCHEOLOGICZNYCH A CECAMI ŚRODOWISKA PRZYRODNICZEGO

Zastosowanie symulacji powoduje, że struktura wewnętrzna poszczególnych drzew decyzyjnych nie może zostać uśredniona. Aby rozpoznać, jakie zmienne są istotne dla wyjaśnienia przestrzennej dyspersji stanowisk archeologicznych, użyto uśrednionej mapy predykcji oraz odchylenia standardowego. W tym celu na nowo wymodelowano drzewo regresyjne, stosując dziewięć tych samych zmiennych niezależnych oraz dwie zmienne zależne jednocześnie: średnią i odchylenie standardowe. W tym celu użyto drzewa regresji wielokrotnej (Kim, Lee 2004), które jest modyfikacją zwykłego drzewa regresyjnego, rozszerzonego o możliwość stosowania wielu zmiennych zależnych. Wewnętrzna struktura tego drzewa posłużyła do wyznaczenia przestrzennych jednostek opisujących zależność pomiędzy parametrami morfometrycznymi a osadnictwem w epoce kamienia.

Model i przestrzenne rozmieszczenie zainteresowania terenami we wczesnej i środkowej epoce kamienia został przedstawiony na rycinach 7 i 8. Klasy o najwyższym zainteresowaniu grup ludzkich znajdują się w prawej gałęzi głównej drzewa decyzyjnego, a ich cechy wskazują na tereny obniżone, położone w dolinach. Obszary wzbudzające największe zainteresowanie (klasa 12) znajdują się najbliżej cieków

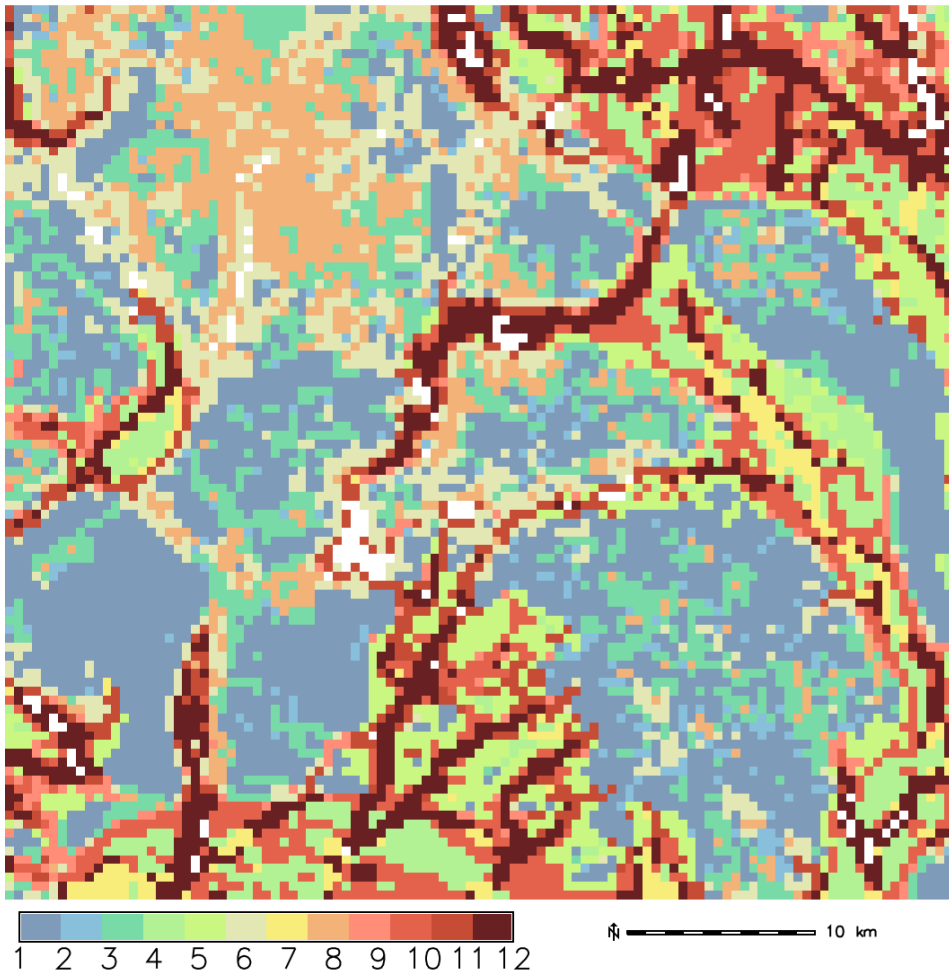


Ryc. 7. Drzewo regresyjne dla końcowego modelu wyjaśniającego zależności pomiędzy intensywnością obecności ludzkiej a cechami przyrodniczymi

Fig. 7. Regressive tree for the final model explaining dependence between intensity of human presence and environmental features

wodnych, pozostałe o dużym zainteresowaniu (klasy 11, 10 i 9) zajmują wyżej położone części doliny, głównie terasy górne.

Druga grupa klas terenu (8, 7, 6, 5 i 4) została określona jako obojętne dla obecności człowieka epoki kamienia. Grupa ta nie jest jednolita. Klasy są rozłożone w obu głównych gałęziach drzewa: 4, 5 i 7 obejmują tereny wyższych partii doliny oraz strefy przejściowe między dolinami i płaskowyżami. Klasy 8 i 6 lokują się w lewej gałęzi drzewa, co oznacza, że obejmują obszary relatywnie wysoko położone. Sama klasa 8 obejmuje rozcięte wzniesienia wzgórz morenowych (wartości najwyższej TEXTURE) w północno-wschodniej części badanego obszaru. Klasa ta może być traktowana jako obszar preferowany-do-obojętne, co oznacza, że na badanym terenie, z wyjątkiem stref dolinnych, zidentyfikowano zupełnie inny obszar, który mógł znaleźć się w polu zainteresowania grup ludzkich epoki kamienia. Klasa 6 obejmuje obszary podobne do 8, ale w mniejszym stopniu porożcinane dolinami. Wskazywałyoby to, że głównym rodzajem aktywności było rybołówstwo, ale ważną rolę odgrywało również łowiectwo – przede wszystkim na obszarach z licznymi dolinkami, do których zapędzano zwierzynę (Bratlund 1999; Spiess 1979).



Ryc. 8. Model zależności pomiędzy intensywnością osadniczą a cechami środowiska przyrodniczego: 1–12 numery klas opisanych w tabeli 5

Fig. 8. Model of dependence between settlement intensity and natural environment features: 1–12 numbers of classes described in table 5

Trzecia grupa, składająca się z klas 1, 2 i 3, została zinterpretowana jako unikana przez społeczności ze starszych okresów epoki kamienia. Wszystkie klasy tezy są w lewej gałęzi drzewa i obejmują głównie płaskowyże morenowe. Klasa 1 ma najniższy odsetek stanowisk w stosunku do zajmowanego obszaru, charakteryzującego się płaską powierzchnią terenu. Pozostałe klasy obejmują tereny o bardziej urozmaiconej rzeźbie terenu (wyższe wartości TEXTURE i RANGE).

WYNIKI I WYMIERNE EFEKTY METODY

W rezultacie przeprowadzonych badań zaproponowano nową metodę analizy lokalizacji stanowisk archeologicznych, wykorzystującą drzewa klasyfikacyjno-regresyjne z użyciem różnych zmiennych przyrodniczych. Wyróżnione klasy: *zawsze pozytywne*, *pozytywne*, *negatywne* i *zawsze negatywne* pozwalają zarówno na planowanie badań archeologicznych, jak również mogą być pomocne przy planowaniu inwestycji. Zaproponowane klasy przetestowano pod względem preferencji lokalizacji ludności ze starszej i środkowej epoki kamienia.

Opracowano szczegółowy model zależności pomiędzy zmiennymi przyrodniczymi a intensywnością aktywności pradziejowej. Wyróżniono 12 klas kontrolowanych przez różne wartości zmiennych. Stwierdzono, że aktywność ludzi epoki kamienia koncentrowała się głównie w pobliżu rzek i dużych jezior, ale zainteresowaniem cieszyły się również urozmaicone obszary wyniesień morenowych silnie porożcinanych formami dolinnymi. Pierwszy rodzaj aktywności związany był zapewne z rybołówstwem, drugi z łowiectwem (Bratlund 1996, 1999). Najmniejszym zainteresowaniem cieszyły się płaskie obszary wysoczyznowe, położone relatywnie daleko od zbiorników wodnych, co zapewne wynikało z niewielkiej aktywności zwierząt łownych w takich regionach (Spiess 1979).

Rezultatem badań jest opracowanie mapy, na której usystematyzowano i przedstawiono w sposób zwarty model obecności ludzkiej w starszej i środkowej epoce kamienia. Wzorec zaprezentowany na mapie jest zgodny z wcześniejszymi spostrzeżeniami odnoszącymi się do terenu Pojezierza Lubuskiego o lokowaniu obozowisk przez schyłkowopaleolityczne i mezolityczne społeczności na terenach położonych w dnach dolin, w pobliżu cieków wodnych, o niewielkim nachyleniu stoków (Kobusiewicz 1969, s. 61–65; Kubiak 1989). Podobne preferencje w lokowaniu obozowisk zaobserwowano także w dolinie Wisły i w zlewisku Wisły oraz Bugo-Narwi (Schild 1975, s. 323). Również w dolinie Niemna stanowiska późnopaleolitycznych łowców-zbieraczy znajdowały się głównie wzdłuż brodów rzecznych, co ułatwiało polowania na przepływającą się zwierzynę (Daugnora, Girininkas 2005). Z terenu Niemiec, z Ahrensburgskiej Doliny Tunelowej znane są liczne stanowiska późnopaleolitycznych łowców (Meindorf, Stellmoor, Borneck, Hopfenbach), usytuowane w pobliżu lub wręcz na krawędzi wąskich dolin o północno-południowym przebiegu (Bratlund 1996). Obozowiska były również zakładane w pobliżu niewielkich zbiorników wodnych ułatwiających nie tylko polowania, lecz także umożliwiających przechowywanie tusz zwierzęcych (Poggenwisch, Meiendorf, Stellmoor) (Bratlund 1991).

Krytycy modelowania predykcyjnego lokalizacji stanowisk łowców-zbieraczy wskazują na kilka istotnych problemów wynikających z traktowania tych społeczności jako pasywnych użytkowników środowiska naturalnego – w opozycji do osadników neolitycznych wpływających aktywnie na jego charakter (Grøn 2010, s. 177).

Zasadniczą wadą modeli predykcyjnych jest założenie stabilności czynników środowiskowych, takich jak: topografia, gleby, hydrologia, wielkość stad, charakter wegetacji. Markery te mogły jednak ulegać zmianie w krótkim okresie (np. zmiana szlaków migracyjnych reniferów), czego model nie jest w stanie uwzględnić (Brandt i in. 1992). Ponadto opieranie się na obserwacjach etnograficznych sugeruje, iż wielkość stada kształtowała się poprzez zabijanie naturalnych drapieżników. W modelach tych nie bierze się również pod uwagę czynników niepragmatycznych, takich jak np. normy kulturowe. Część badaczy podkreślając rolę relacji społecznych, naturalnych i kulturowych obiektów, a także norm społecznych w kreowaniu przestrzeni, twierdzi, że przestrzeń jest współcześnie rozumiana w zbyt zdehumanizowany sposób (Tilley 1994, s. 17; 2008, s. 28)³. Należy jednak podkreślić, że model jedynie wskazuje obszary spodziewanej obecności śladów działalności człowieka, które powinny zostać zweryfikowane w wyniku badań powierzchniowych, sondażowych lub wykopaliskowych.

Zastosowanie modelu prognostycznego przez J. Jasiewicza i I. Hildebrandt-Radke pokazało, że dyspersja stanowisk w południowej części Niziny Wielkopolskiej, na Równinie Kościańskiej, wykazuje nielosowe relacje w stosunku do elementów rzeźby, a co więcej, że zmieniały się one w czasie (Hildebrandt-Radke 2013, s. 38–57; Jasiewicz, Hildebrandt-Radke 2009a, 2009b).

Mimo wielu ograniczeń modelowanie przestrzenne (Stach 2011, s. 94) z zastosowaniem metody drzew klasyfikacyjno-regresyjnych jest skutecznym narzędziem w definiowaniu środowiskowych uwarunkowań lokalizacji stanowisk archeologicznych oraz prognozowaniu ich potencjalnej lokalizacji. Nie jest natomiast i nie może być jedynym źródłem wiedzy na temat zachowań pradziejowych społeczności ludzkich.

BIBLIOGRAFIA

- Bower J., Kobusiewicz M.
 2002 *Comparative Study of Prehistoric Foragers in Europe and North America. Cultural Responses of the End of the Ice Age*. New York: The Edwin Mellen Press.
- Brandt R., Groenwoudt B.J., Kvamme K.L.
 1992 An Experiment in Archaeological Site Location: modelling in the Netherlands using GIS Techniques. *World Archaeology*, 12(1–3), s. 19–202.
- Bratlund B.
 1991 Die spätglazialen „Opfertiere“ von Meiendorf und Stellmoor, Kreis Stormarn. Neue Ansätze zur Interpretation alter Funde. *Offa*, 48, s. 41–73.

³ O. Grøn (2010) podaje przykład kanadyjskiego plemienia Mistassini Cree, któremu reguły kulturowe zakazują picia wody powstałej z rozpuszczonego lodu i śniegu – w opozycji do syberyjskich Evenków preferujących ten sposób pozyskiwania wody i lokujących swoje zimowe siedziby nawet do 5 km od zbiorników wodnych.

- Bratlund B.
1996 Hunting Strategies in the Late Glacial of Northern Europe. A Survey of the Faunal Evidence. *Journal of World Prehistory*, 10(1), s. 1–48.
- Bratlund B.
1999 A Survey of the Ahrensburgian Faunal Assemblage of Stellmoor. W: S.K. Kozłowski, J. Gurba, L. Zalizniak, *Tanged Point Cultures in Europe. Read at the International Archaeological Symposium Lublin, September, 13-16, 1993* (Lubelskie Materiały Archeologiczne, t. 13, s. 47–59). Lublin: Maria Curie-Skłodowska University Press.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J.
1984 *Classification and Regression Trees*. New York: Chapman & Hall.
- Chmal R.
2003 *Objaśnienia do Szczegółowej Mapy Geologicznej Polski. Arkusz Toporów (501)*. Warszawa: Państwowy Instytut Geologiczny.
- Cook E.F., Goldman L.
1984 Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *Journal of Chronic Diseases*, 37, s. 9–10.
- Daugnora L., Girininkas A.
2005 Šiaurės elnių keliai ir jų paplitimas lietuvoje vėlyvajame paleolite. *Lietuvos Archeologija*, 29, s. 119–132.
- Duke C.
2003 Quantifying Palaeolithic landscapes: computer approaches to terrain analysis and visualisation. W: M. Doerr, A. Sarris (red.), *The Digital Heritage of Archaeology* (s. 139–146). Athens: Hellenic Ministry of Culture.
- Estrabrooks A., Taeho J., Japkowicz N.
2004 A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*, 20(1), s. 18–36.
- Fayyad U., Piatetsky-Shapiro G., Smyth P.
1996 From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), s. 37–54. [Pobrane z: <http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1230>].
- Garcia E.A., Haibo H.
2009 Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), s. 1263–1284.
- Goldman L., Weinberg M., Weisberg M., Olshen R., Cook E.F., Sargent R.K., Lamas G.A., Dennis C., Wilson C., Deckelbaum L., Fineberg H., Stiratelli R.
1982 A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *The New England Journal of Medicine*, 307(10), s. 588–596.
- Grøn O.
2010 Our grandfather sent the elk – some problems for hunter-gatherer predictive modelling. *Quartär*, 59, s. 175–188.
- Hildebrndt-Radke I.
2013 *Pradziejowa i wczesnohistoryczna antropopresja i jej zapis w środowisku na przykładzie regionu środkowej Obry (Wielkopolska)*. Poznań: Bogucki Wydawnictwo Naukowe.
- Japkowicz N.
2000 *Learning from Imbalanced Data Sets: a Comparison of Various Strategies*. [Pobrane z: <http://www.aaai.org/Papers/Workshops/2000/WS-00-05/WS00-05-003.pdf>].
- Jankowska D., Rozbiegalski P.
2012 Między północą a południem. Osadnictwo schyłkowopaleolityczne i mezolityczne w dolnym biegu Warty. *Folia Praehistorica Posnaniensia*, 17, s. 139–148.

- Jaskanis D.
1992 Polish national record of archaeological sites. General outline. W: C.V. Larsen (red.), *Sites and Monuments. National Archaeological Records* (s. 81–87). Copenhagen: National Museum of Denmark.
- Jasiewicz J., Hildebrandt-Radke I.
2009a Using multivariate statistics and fuzzy logic system to analyse settlement preferences in lowland areas of the temperate zone: an example from the Polish Lowlands. *Journal of Archaeological Science*, 36, s. 2096–2107.
- Jasiewicz J., Hildebrandt-Radke I.
2009b Zastosowanie oprogramowania OPEN SOURCE GIS do oszacowania wpływu parametrów morfometrycznych terenu na rozwój osadnictwa pradziejowego, na przykładzie Równiny Kościańskiej. W: A. Kostrzewski, R. Paluszkiewicz (red.), *Geneza, litologia i stratygrafia utworów czwartorzędowych (Seria Geografia 88, tom V, s. 155–167)*. Poznań: Wydawnictwo Naukowe UAM.
- Jasiewicz J., Stepinski T.
2013 Geomorphons – a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182, s. 147–156.
- Kabaciński J., Sobkowiak-Tabaka I.
2010 Osadnictwo późnopaleolityczne. W: J. Kabaciński, I. Sobkowiak-Tabaka (red.), *Materiały do wczesnych pradziejów Zachodniej Wielkopolski. Osadnictwo późnopaleolityczne i mezolityczne na stanowisku 7 w Ośnie Lubuskim. Ratownicze badania archeologiczne Instytutu Archeologii i Etnologii PAN, Oddział w Poznaniu II* (s. 17–60). Poznań: Instytut Archeologii i Etnologii PAN.
- Kabaciński J., Sobkowiak-Tabaka I.
2011 Paleolit. W: J. Kabaciński, I. Sobkowiak-Tabaka (red.), *Materiały do wczesnych pradziejów Zachodniej Wielkopolski. Osadnictwo pradziejowe i wczesnośredniowieczne w Lubrzy. Ratownicze badania archeologiczne Instytutu Archeologii i Etnologii PAN, Oddział w Poznaniu III* (s. 195–262). Poznań: Instytut Archeologii i Etnologii PAN.
- Kabaciński J., Sobkowiak-Tabaka I.
2013 *Osadnictwo społeczności kultury hamburskiej na Pojezierzu Lubuskim. Ratownicze badania Archeologiczne Instytutu Archeologii i Etnologii PAN, Ośrodek Studiów Pradziejowych i Średniowiecznych V*. Poznań: Instytut Archeologii i Etnologii PAN.
- Kim S.J., Lee K.B.
2003 Constructing Decision Trees with Multiple Response Variables. *International Journal of Management and Decision Making*, 4, s. 337–353.
- Kobusiewicz M.
1999 *Ludy łowiecko-zbierackie północno-zachodniej Polski*. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk.
- Kondracki J.
2009 *Geografia regionalna Polski*. Warszawa: Wydawnictwo Naukowe PWN.
- Konopka M. (red.)
1981 *Archeologiczne Zdjęcie Polski* (Biblioteka Muzealnictwa i Ochrony Zabytków, seria B, 46). Warszawa: Ministerstwo Kultury i Sztuki.
- Kubiak L.
1989 *Środowiska osadnicze ludności łowiecko-zbierackiej w zachodniej części Niziny Wielkopolskiej w późnym paleolicie i mezolicie*. [Maszynopis pracy magisterskiej, Poznań].
- Marks L., Ber A., Gogolek W., Piotrowska K.
2006 *Geologic Map of Poland 1:500 000 with planatory Text*. Warszawa: Państwowy Instytut Geologiczny.

- Misiewicz K.
2006 *Geofizyka archeologiczna*. Warszawa: Instytut Archeologii i Etnologii PAN.
- Okuniewska-Nowaczyk I.
2005 Stan badań palinologicznych w południowo-zachodniej Wielkopolsce. *Fontes Archaeologici Posnanienses*, 41, s. 9–13.
- Okuniewska-Nowaczyk I., Sobkowiak-Tabaka I.
2014 Palynological and archaeological studies of the Late Vistulian and Early Holocene at the Lubuskie Lake District (Western Poland). *Studia Quaternaria*, 31(1), s. 39–50.
- Prinke A.
2002 Introducing information technology to archaeological resource management: towards a GIS-based SMR of Midwestern Poland. W: L. Garcia Sanjuan, D. Wheatley (red.), *Mapping the Future of the Past. Managing the Spatial Dimension of the European Archaeological Resource* (s. 85–96). Sevilla: Universidad de Sevilla.
- Sobkowiak-Tabaka I.
2014 The recolonisation of the Polish Lowland – new ideas and discoveries. W: F. Riede, M. Tallaavaara (red.), *Lateglacial and Postglacial Pioneers in Northern Europe* (British Archaeological Reports, International Series 2599, s. 53–65). Oxford: Archaeopress.
- Sobkowiak-Tabaka I., Okuniewska-Nowaczyk I., Ratajczak-Szczerba M.
[w druku] *Lateglacial human occupation in the Lubrza region (Łagów Lake District, Western Poland)*.
- Spiess A.F.
1979 *Reindeer and Caribou Hunters. An Archaeological Study*. New York: Academic Press.
- Stach A.
2011 Statystyka przestrzenna w archeologii: opis, analiza i prognozowanie. W: J. Jasiewicz, M. Lutyńska, M. Rzeszewski, M. Szmyt, M. Makohonienko (red.), *Metody geoinformacyjne w badaniach archeologicznych. Środowisko i kultura 9* (s. 91–95). Poznań: Bogucki Wydawnictwo Naukowe.
- Strahler A.N.
1952 Hypsometric (area-altitude) analysis of erosional topography. *Bulletin of the Geological Society of America*, 63, s. 1117–1142.
- Vanacker V., Govers G., Van Peer G., Verbeek C., Desmet J., Reyniers J.
2001 Using Monte Carlo simulation for the environmental analysis of small archaeological datasets, with the Mesolithic in Northeast Belgium as a case study. *Journal of Archaeological Science*, 28(6), s. 661–669.
- Williams A.N., Ulm S., Cook A.R., Langley M.C., Collard M.
2013 Human refugia in Australia during the Last Glacial Maximum and Terminal Pleistocene: a geospatial analysis of the 25–12 ka Australian archaeological record. *Journal of Archaeological Science*, 40(12), s. 4612–4625.
- Tilley Ch.
1994 *A Phenomenology of Landscapes. Places, Paths and Monuments*. Oxford: Berg 3PL.
- Tilley Ch.
2010 *Interpreting Landscapes. Geologies, Topographies, Identities. Explorations in Landscape Phenomenology 3*. Walnut Creek: Left Coast Press Inc.
- Zhang H.
1998 Classification Trees for Multiple Binary Responses. *Journal of the American Statistical Association*, 93, s. 180–193.

ANALYSIS OF ENVIRONMENTAL VARIABLES OF PALAEOLITHIC
AND MESOLITHIC SETTLEMENT LOCATION IN POJEZIERZE LUBUSKIE.
APPLICATION OF CLASSIFICATION AND REGRESSION TREES

S u m m a r y

Recognition of the variables influencing location of Stone Age sites is particularly important in regard to lowlands, lacking in easily recognisable objects that might attract human presence. Additionally late-glacial landscapes are hardly accessible for archaeological prospection due to frequent meadows, woodlands or peat bogs in the area. Preliminary analyses of the distribution of archaeological sites within a part of Pojezierze Lubuskie have shown that over 50% of the area is not available to traditional field walking, and non-invasive methods, such as electrical resistivity imaging, magnetic or electromagnetic methods, are not appropriate tools for registration of relics of human activity, especially from Palaeolithic and Mesolithic periods. Therefore, spatial analysis, applying statistical modelling, becomes the only alternative to invasive methods of potential sites location.

Analysis of environmental variables to predict the settlement location has been based on the application of methods associated with knowledge discovery in databases, which is based on data mining (exploration of data). The basic method employed for classification and analysis has been classification-regression tree modeling (CART). This is one of the most popular machine learning algorithms. The main advantage of using CART is that obtained model is not based on any statistical assumptions adopted *a priori*.

In the prediction of sites location the used data have been obtained from the Archaeological Record Cards of sites together with derivatives of the Digital Terrain Elevation model (DEM 30 m).

While analysing location of Palaeolithic and Mesolithic sites in Pojezierze Lubuskie, nine parameters were regarded, such as RELBASE (relative height above the surface), RANGE (difference between the highest and the lowest elevations within a grid cell of a raster created in Digital Elevation Model – DETD), FLATNESS (value from 0 to 1 for flat or slightly sloped terrains, less than 4°, in a particular grid cell), TEXTURE (summarised distance from an indentation, e.g. a valley, in a particular grid cell of analysed raster), LOWLAND – value of terrains below an average in a grid cell, WETNESS – topographic (TCI) or wetness index, DISTWATER – counted as a minimal hydrological distance along a watercourse from a particular grid cell to the nearest lake or river, FORMS – categorical variable, GEOL – lithology of a surface.

As a result of the study a new method of predicting location of archaeological sites has been suggested, applying classification and regression trees with different natural variables. Distinguished classes: always positive, positive, negative, and always negative permit planning both archaeological research and investments. The proposed classes were tested, in terms of preferences of Palaeolithic and Mesolithic people with regard to site location.

A detailed model of the relationship between natural variables and intensity of prehistoric activity has been developed, with division into twelve classes controlled by different variables. It has been ascertained that activity of the Stone Age people concentrated mainly nearby rivers and large lakes, although varied elevated areas of moraine forms, strongly incised by river valleys, were also

valued. The least appreciated were flat plateaus, located relatively far from the water reservoirs, which was probably due to the low activity of game animals in such regions.

The result of the research is a map that systematises and presents, in a coherent manner, a model of human presence in the Palaeolithic and Mesolithic periods.

Despite its numerous limitations, spatial modelling, using a method of classification and regression trees, appears to be an effective tool in defining the environmental variables of the location of archaeological sites and predicting their potential location. However, it is not, and might not be, the only source of knowledge about the behaviour of prehistoric communities.

Translated by Lucyna Leśniak