



## Statistische Vorhersage in der Archäologie mithilfe spezieller binärer Funktionen

*Olga Wälder, Dresden*

### Kurzfassung

Bei einer kompetenten Anwendung statistischer Verfahren muss stets darauf geachtet werden, dass solche Grundannahmen wie Unabhängigkeit sowie Verteilungsannahmen und Ursache-Wirkungs-Relationen für die in die Modellierung eingehenden Variablen erfüllt sind. In den Geowissenschaften ist dies nur sehr schwer oder fast unmöglich zu gewährleisten. In dieser Arbeit wird ein neues Vorhersageverfahren präsentiert, das die mathematische Modellierung nicht eindeutiger Ursache-Wirkungs-Relationen anhand spezieller binärer Funktionen nutzt. Seine Anwendung wird anhand einer archäologischen Fallstudie präsentiert.

**Schlüsselwörter:** Statistik, Archäologie, binäre Funktionen

### Abstract

A fair application of statistical methods requires the fulfilling of such basic properties as independency, distribution rules, action-reaction-relations for the variables which are taken into modelling. This is very difficult or nearly impossible in the geosciences. In this paper a new approach is presented that allows mathematical modelling of non explicit action-reaction-relations. Our method is based on using special binary functions for statistical predictions. An application of our method in archaeology is presented.

**Keywords** statistic, archaeology, binary functions

### 1. Einleitung

Statistische Verfahren basieren stets auf bestimmten Annahmen. So müssen die unabhängigen Variablen bei diversen Regressionsansätzen tatsächlich unabhängig sein und die abhängige Variable ist dabei irgendeine Funktion, die von diesen abhängt. Jeder explizite funktionale Zusammenhang der Form  $Y = f(X_1, \dots, X_n)$  spiegelt eine vorausgesetzte Ursache-Wirkungs-Relation zwischen Variablen wider. Bei vielen Anwendungen in den Geowissenschaften und auch der Archäologie können die erforderlichen statistischen Modellannahmen nicht gewährleistet werden. Unter anderem kann in der Archäologie nicht davon ausgegangen werden, dass die Fundorte von früheren Siedlungen in einer klaren Ursache-Wirkungs-Relation zu solchen geologischen Faktoren wie Bodenbeschaffenheit, Gewässernähe und Geländeneigung stehen. Dies liegt vor allem daran, dass die meisten dieser Faktoren im Verlauf der menschlichen Geschichte anthropogen beeinflusst wurden.

Dennoch versuchen archäologische Vorhersagemodelle, den frühzeitlichen Entscheidungsprozess unserer Vorfahren für die Festlegung eines Siedlungsplatzes nachzuvollziehen. Die zwar durchaus logisch klingende, aber den möglichen Entscheidungsprozess nur einseitig modellie-

rende Annahme ist dabei, dass die Siedlungswahl teilweise durch Umweltfaktoren beeinflusst wurde, siehe Posluschny (2002). In den USA wurden die ersten archäologischen Prognosemodelle (Predictive Models) für statistische Landverwaltungsprojekte ab den 1970er Jahren entwickelt. Ab den 1980er Jahre wurden sie in Geographische Informationssysteme (GIS) implementiert und seitdem auch weiterentwickelt. In erster Linie geht es bei diesen Modellen um Kulturerbeschutz und -pflege. Nach Groenewoudt et al. (1994) in van Leusen und Kamermans (2005) wurde ein Drittel der bis 1950 nicht erfassten archäologisch interessanten Stellen bis heute bereits zerstört.

In Jackenkroll (2008) wurden drei verschiedene Vorhersagemodelle anhand einer Fallstudie miteinander verglichen. Dort erfolgt eine statistische Auswertung des Zusammenhanges zwischen Siedlungsfundstellen und drei Umweltfaktoren, nämlich Entfernung zu einer Naturraumgrenze, Entfernung zum Gewässer und Exposition. Eine daraus abgeleitete Prognose soll zu einer Wahrscheinlichkeitskarte für weitere Siedlungen führen. Dabei wird erneut eine eindeutige Ursache-Wirkungs-Relation zwischen Variablen unterstellt, die nach der kritischen Meinung von Jackenkroll (2008) zu keinen ausagekräftigen Ergebnissen geführt hat.

Die Anwendung geowissenschaftlicher Methoden und Konzepte auf archäologische Fragestellungen ist auch unter dem Oberbegriff „Geoarchäologie“ bekannt. Das Ziel der Geoarchäologie ist die Archäoprognose. Diese betrifft in erster Linie das Erkennen, Datieren und Erklären von Befunden, vgl. Brückner und Gerlach (2006). Ergänzt wird diese Wissenschaft durch die räumliche Analysemethoden sowie das Interesse am „Urzustand“ der Landschaft, vgl. Sabel (1982). Viele archäologische Vorhersagemodelle sind im statistischen Sinne zumindest unausgereift und in vielen Fällen komplett unbrauchbar. In dieser Arbeit wird ein neues statistisches Vorhersagemodell präsentiert, das vor allem auf die Annahme einer eindeutigen Ursache-Wirkungs-Relation zwischen den Variablen verzichtet. Wir haben unser Vorhersagemodell als LOG-Modell bezeichnet, weil es sich aus mathematischer Sicht auf eine spezielle Verwendung von elementaren logisch-theoretischen binären Funktionen stützt.

Die Anwendung des Verfahrens erfolgt auf der Basis von Daten aus Saile (1998) und Jackenkroll (2008) und wurde mithilfe des Programmtools MATLAB numerisch umgesetzt. Seine Implementierung in GIS ist geplant. Bei der Anwendung geht es um die Erfassung, Analyse und Modellierung der Zusammenhänge zwischen einigen geologischen Input-Variablen und der Verteilung der Fundorte von früheren Siedlungen im Bereich der Wetterau (Hessen). Eine Vorhersage der Standorte (Wahrscheinlichkeitskarte) der möglichen, noch nicht entdeckten Fundorte wird zudem angestrebt. Diese Arbeit ist in folgende Arbeitsschritte gegliedert:

- 1) Primäre Datenaufbereitung
- 2) Beschreibung zweier LOG-Modelle
- 3) Auswertung der Ergebnisse der Fallstudie „Wetterau“
- 4) Vorhersage anhand der Modelle
- 5) Kritische Diskussion der Problematik

## 2. Beschreibung des Untersuchungsgebietes und des Datenmaterials

Das Untersuchungsgebiet liegt in der Wetterau im Bundesland Hessen. Die Wetterau ist eine vom Taunus im Westen, im Norden vom Giessener Landrücken und dem Durchbruch in das Lahnetal, im Osten durch den Vogelsberg und im Süden durch den Frankfurter Horst und das Maintal begrenzte Beckenlandschaft. Das Gebiet hat eine Ausdehnung von etwa 20 km in West-Ost und 40

km in Nord-Süd-Richtung. Gebildet hat sich die Region vor ca. 300 Millionen Jahren.

Die nördliche Wetterau ist durch ihre trockenen klimatischen Bedingungen am Anfang des Boreals durch Kiefern und Haselsträucher geprägt. Mit der anhaltenden Erwärmung wurden sie durch Linden, Eichen und Ulmen ergänzt. Der Raum war fast vollständig bewaldet. Diese Situation veränderte sich zunehmend unter anthropogenem Einfluss. Von den ersten inselartigen Rodungen kam es später zur Beweidung der Wälder. Die gute Bodenqualität führte zur extensiven Landwirtschaft. An Orten mit günstigen Standortfaktoren bildeten sich Siedlungscluster heraus. Saile (1998) stellte dabei die Präferenz von Siedlungsorten an Naturraumgrenzen und der Nähe zum Gewässer fest.

Die Siedlungsfundorte der folgenden drei Kulturepochen werden in dieser Arbeit untersucht, vgl. auch Jakenkroll (2008): Die Epoche der Bandkeramiker, die Epoche der Urnenfelder und die Epoche der römischen Kaiserzeit. Die Kulturepoche der Bandkeramiker entspricht der Zeit von ca. 5000 v. Chr. und zeichnete sich durch eine starke Homogenität aus.

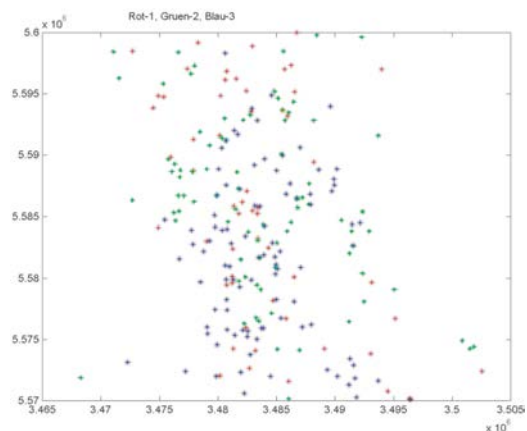


Abb. 1: Entdeckte Siedlungsstandorte in den Epochen der Bandkeramiker (rot), der Urnenfelder (grün) und der römischen Kaiserzeit (blau).

Die Kultur hebt sich in Bauweise, Handwerk und kultischem Verhalten von der mittelsteinzeitlichen Epoche der Jäger, Fischer und Sammler ab, siehe Zimmermann (2002). Die zur diesen Epoche gehörigen Siedlungsfundorte sind in Abbildung 1 rot gekennzeichnet.

Die Urnenfelderkultur entspricht der späten Bronzezeit. Älteste Fundstellen lassen sich auf

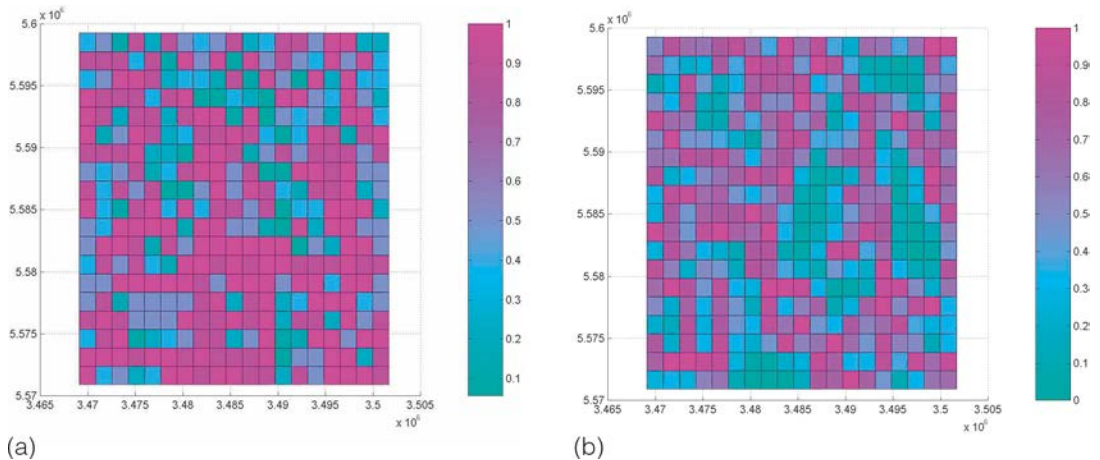


Abb. 2: Gemittelte Input-Variablen: Nahe zur Naturraumgrenze  $B$  (a) und Nahe zum Gewasser  $W$  (b) auf einem  $20 \times 20$  Gitter.

1300 v. Chr. datieren. Es verbreitete sich die Nutzung des Pfluges ber Mitteleuropa und es wurde Roggen und Hafer zusatzlich zu den anderen Ackerfrchten angebaut. Ahnlich wie die Bandkeramiker bevorzugten auch die Urnenfelder bei ihrer Standortwahl Grenzstandorte zwischen verschiedenen Bodentypen (d.h. die Nahe zur Naturraumgrenze), vgl. Saile (1998). Die zur Epoche der Urnenfelder gehrigen Siedlungsfundorte sind in Abbildung 1 grn markiert.

Die Epoche der rmischen Kaiserzeit begann um ca. 100 v. Chr. und dauerte bis 300 n. Chr. Nur ein Teil der Wetterau stand unter der Herrschaft der Rmer. Die Rmer pragten die Siedlungsstruktur stark durch den Grenzwall „Limes“. Im Vorland wurden Kastelle errichtet. In ihrer Nahe breiteten sich Lagedrfer aus. Ausweitung der Landwirtschaft fand statt. Die zur diesen Epoche gehrigen Siedlungsfundorte sind in Abbildung 1 blau gekennzeichnet.

Die Klassifizierung der Bodentypen wird nach der hessischen Bodenkarte im Mastab 1:50 000 durchgefhrt. Dabei mssen einige Korrekturen vorgenommen werden, weil diese Information nicht mit den Gegebenheiten von vor ber 1000 Jahren gleichgesetzt werden kann, vgl. Jackenkroll (2008). Zum Definitionsproblem eines „guten Bodens“ verweist Sabel (1982) auf die hohe Bedeutung der Drainage fr die frhen Ackerbauern. Dadurch wurde die Bearbeitung des Bodens erleichtert, aber auch sein ursprnglicher Zustand verandert. Die Betrachtung der Nahe zur Naturraumgrenze anstatt des Parameters „Boden“ raumt einige diesbezgliche Unklarheiten aus. Fr diesen Parameter wurden bei der

spateren Modellierung 18 Werte zwischen 0 (weit) und 1 (nah) vergeben. Ahnlich wird der zweite Parameter „Nahe zum Gewasser“ klassifiziert: Er kann 18 Werte von 0 (optimale Lage) bis 1 (zu weit) annehmen. Jackenkroll (2008) benutzt den zusatzlichen Parameter „Exposition“, auf den wir in dieser Arbeit wegen seiner relativ geringen Bedeutung verzichten.

### 3. Primare Datenaufbereitung

Die folgenden Variablen sollen als Input-Variablen bezeichnet werden: Nahe zur Naturraumgrenze  $B$  und Nahe zum Gewasser  $W$ .

Wegen einer gewissen Unsicherheit bei den Werten der Input-Variablen wird im Weiteren nur mit ihren Mittelwerten gearbeitet. Das untersuchte Gebiet wird in  $20 \times 20$  gleichgroe Quadrate eingeteilt, vgl. Abbildung 2. In jeder Masche werden die Mittelwerte der entsprechenden Input-Variablen berechnet. Den Zentren dieser Quadrate werden dann diese Mittelwerte zugeordnet (daher das endgltige  $19 \times 19$  – Gitter in der Abbildung). Auf diesem Wege entstehen zwei Gitter mit den Werten fr die Mittelwerte der Parameter  $B$  und  $W$ , die in die spatere Analyse als Input-Variablen eingehen, siehe Abbildung 2.

Abbildung 3 zeigt den so genannten Favorisierungsindex. Dieser Index stellt eine zusatzliche Variable dar. Diese spiegelt die Favorisierung entsprechender Bereiche anhand der relativen Anzahl aller entdeckten Siedlungen in jeder Masche wieder.

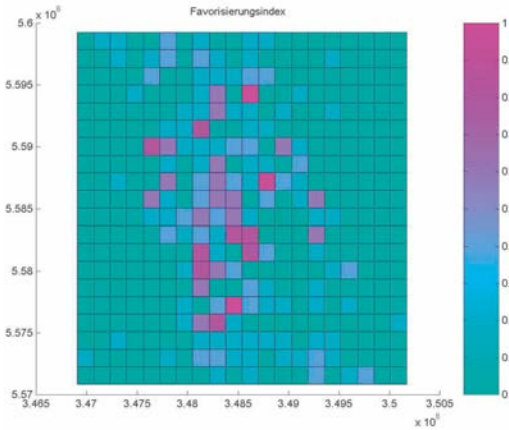


Abb. 3: Favorisierungsindex  $Y_3$  für alle entdeckten Siedlungsstandorte.

Diese relativen Anzahlen von Siedlungen werden dabei auf das Intervall  $[0,1]$  transformiert (durch ihr Maximum geteilt). Die berechneten Werte werden den Zentren von Maschen zugeordnet. Wir bezeichnen das entsprechende Gitter mit  $Y_3$ . Ähnlich werden auch Favorisierungsindizes für die Fundorte der Siedlungen der Bandkeramiker  $Y_1$  sowie für die Fundorte der Siedlungen der Bandkeramiker und der Urnenfelder  $Y_2$  erstellt, siehe Abbildung 4.

4. Zwei LOG-Modelle

Zunächst möchten wir die zwei bekannten logisch-theoretischen binären Funktionen „AND“ ( $x_1 \cdot x_2$ )

und „OR“ ( $x_1 + x_2$ ) in der üblichen tabellarischen Form präsentieren:

$x_1$	$x_2$	$x_1 \cdot x_2 = \min(x_1, x_2)$	$x_1 + x_2 = \max(x_1, x_2)$
0	0	0	0
1	0	0	1
0	1	0	1
1	1	1	1

Wir werden zwei weitere binäre Funktionen verwenden, nämlich

$$F_1(x_1, x_2) = (x_1 \cdot x_2) + (x_1^C \cdot x_2^C)$$

und

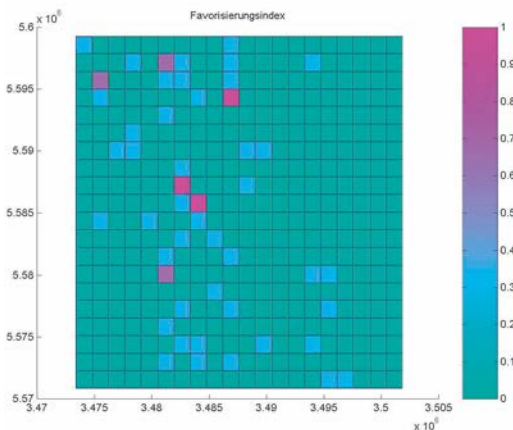
$$F_2(x_1, x_2, x_3) = (x_1 + x_2) \cdot x_3 + (x_1^C \cdot x_2^C) \cdot x_3^C$$

Mit dem Symbol „C“ wird dabei das Komplement von x (d.h.  $1-x$ ) bezeichnet. Der Wertebereich der ersten Funktion  $F_1(x_1, x_2)$  entspricht:

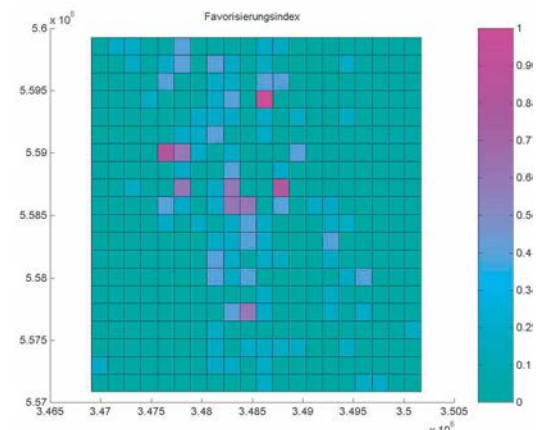
$x_1$	$x_2$	$(x_1 \cdot x_2) + (x_1^C \cdot x_2^C)$
0	0	1
1	0	0
0	1	0
1	1	1

Diese Funktion wird in der weiteren Modellierung einen starken logisch-theoretischen Zusammenhang zwischen Input-Variablen beschreiben. Ihre Erweiterung für den Fall von drei Variablen entspricht nun der Funktion

$$F_1(x_1, x_2, x_3) = (x_1 \cdot x_2 \cdot x_3) + (x_1^C \cdot x_2^C \cdot x_3^C)$$



(a)



(b)

Abb. 4: Favorisierungsindizes  $V_1$  (a) und  $Y_2$  (b).

Diese Funktion erhalt den Wert Eins im Fall  $x_1 = x_2 = x_3 = 0$  oder  $x_1 = x_2 = x_3 = 1$ . Fur andere Variablenkombinationen nimmt diese Funktion stets den Wert 0 an.

Die zweite Funktion

$F_2(x_1, x_2, x_3) = (x_1 + x_2) \cdot x_3 + (x_1^C \cdot x_2^C) \cdot x_3^C$  beschreibt einen schwachen logisch-theoretischen Zusammenhang zwischen Input-Variablen. Sie nimmt den Wert 0 in dem Fall, wenn beide Summanden gleich 0 sind.

Die Verallgemeinerung beider Funktionen fur den Fall mehrerer Variablen ist offensichtlich.

#### 4.1 Ein erstes LOG-Modell fur den starken logisch-theoretischen Zusammenhang

Hier wird vom folgenden impliziten Zusammenhang ausgegangen:

$$(B \cdot W \cdot Y_k \cdot Y_{k+1}) + (B^C \cdot W^C \cdot Y_k^C \cdot Y_{k+1}^C) = 1 \quad (1)$$

$k = 1, 2$

Die Relation (1) modelliert den idealisierten Zusammenhang zwischen den Siedlungsfundorten einer Epoche, den fruheren Siedlungsfundorten sowie den Input-Variablen fur jede Masche des Wertegitters. Es wird also unterstellt, dass nur die folgenden Ergebnisse in derselben Masche auftreten konnen:

- a) „Optimale Lage“, fruhere Siedlungen und neue Siedlungen oder
- b) „schlechte Lage“, keine Siedlungen und keine neue Siedlungen.

Die Bezeichnungen „optimale Lage“ sowie „schlechte Lage“ beziehen sich auf die Nahe zur Naturraumgrenze und die Nahe zum Gewasser, siehe die Klassifikation in Abbildung 2. Beide Parameter sollen dabei „optimal“ sein.

#### 4.2 Ein zweites LOG-Modell fur den schwachen logisch-theoretischen Zusammenhang

Eine abgeschwachte Version des Modells (1) fuhrt zu:

$$(B + W + Y_k) \cdot Y_{k+1} + (B^C + W^C + Y_k^C) \cdot Y_{k+1}^C = 1 \quad (2)$$

$k = 1, 2$

Die Relation (2) modelliert den idealisierten Zusammenhang zwischen Siedlungsfundorten der nachfolgenden Zeitepoche, den Input-Variablen sowie den fruheren Siedlungsfundorten fur jede Masche des Wertegitters. Die folgenden Ergebnisse finden in derselben Masche statt:

- a) „Optimale Lage“ oder fruhere Siedlungen und neue Siedlungen oder

- b) „schlechte Lage“ oder keine fruhere Siedlungen und keine neuen Siedlungen.

Die Bezeichnungen „optimale Lage“ sowie „schlechte Lage“ beziehen sich erneut auf die Nahe zur Naturraumgrenze und die Nahe zum Gewasser, siehe ihre Klassifikation in Abbildung 2. Nun soll in diesem Fall mindestens einer der Parameter „optimal“ sein.

#### 4.3 Binarisierung der Variablen und statistische Testgroen

Sowohl die Werte der Input-Variablen als auch die Werte der Favorisierungsindizes werden binarisiert. Dafur wird ein Parameter  $0 < \varepsilon < 1$  eingefuhrt, der als der Schwellenwert dienen soll, d.h.

$$V^* = \begin{cases} 1, & V \geq \varepsilon \\ 0, & \text{sonst} \end{cases}$$

$$V = B, W, Y_k, \quad k = 1, 2, 3 \quad (3)$$

Es werden verschiedene Werte fur den Parameter  $\varepsilon$  bei der Fallstudie getestet. Mit den binarisierten Variablen wird die Hypothese uber die Modellannahme (1) oder (2) gepruft.

Fur die statistische Analyse werden fur die  $k$ -te Epoche die folgenden Testgroen  $T_k, k = 2, 3$  eingefuhrt:

$$T_k = \frac{\text{Anzahl der Maschen mit } F_i = 1}{\text{Anzahl der Maschen insgesamt}}$$

$k = 2, 3$   
 $i = 1, 2$  (4)

wobei fur das erste Modell aus 4.1

$$F_1 = (B^* \cdot W^* \cdot Y_{k-1}^*) \cdot Y_k^* + (B^{*C} \cdot W^{*C} \cdot Y_{k-1}^{*C}) \cdot Y_k^{*C}$$

$k = 2, 3$

und fur das zweite Modell aus 4.2

$$F_1 = (B^* + W^* + Y_{k-1}^*) \cdot Y_k^* + (B^{*C} + W^{*C} + Y_{k-1}^{*C}) \cdot Y_k^{*C}$$

$k = 2, 3$

benutzt werden.

Die Hypothese  $H_0$  „Zwischen der Epoche  $k$  und der Epoche  $k-1$  besteht ein logisch-theoretischer Zusammenhang nach (1) bzw. nach (2)“ wird auf einem Signifikanzniveau  $\alpha$  abgelehnt, falls  $T_k < 1 - \alpha, k = 2, 3$  aus (4) fur  $i=1$  bzw.  $i=2$ . Die ublichen Werte fur  $\alpha$  sind dabei 0.01, 0.05, 0.1.

#### 4.4 Vorhersage mit den obigen Modellen

Unter der Voraussetzung, dass einer der Zusammenhange (1) oder (2) zwischen den Variablen angenommen werden kann, wird im nachsten



Schritt eine archäologische Vorhersage durchgeführt. Die in (5)–(7) benutzten Variablen sind nicht binarisiert.

Die folgenden Schätzer  $\hat{Y}_{k+1}$ ,  $k = 1, 2, \dots$  für die entsprechenden Favorisierungsindizes auf dem  $20 \times 20$  – Gitter können berechnet werden:

$$\begin{aligned} \hat{Y}_{k+1} &= f_{i,k} \\ i &= 1, 2 \\ k &= 1, 2 \end{aligned} \quad (5)$$

wobei für das 1. Modell in 4.1

$$f_{1,k} = \min(\min(B, W), Y_k), \quad k = 1, 2, \dots \quad (6)$$

und für das 2. Modell in 4.2

$$f_{2,k} = \min(\max(B, W), Y_k), \quad k = 1, 2, \dots \quad (7)$$

verwendet werden.

Die Güte dieser Vorhersagemethode kann dann wie folgt geprüft werden. Die bisher entdeckten, also bekannten Fundorte in den Epochen der Urnenfelder und der römischen Kaiserzeit werden laut (5) vorhergesagt. Die Schätzer werden sowohl für die Epoche der Urnenfelder ( $k = 1$ ), als auch für die Epoche der römischen Kaiserzeit ( $k = 2$ ) nach (5) ermittelt. Die vorhergesagten Werte  $\hat{Y}_2, \hat{Y}_3$  werden mit den gegebenen  $Y_2, Y_3$  verglichen, vgl. Abbildungen 2 und 3. Hierzu wird die mittlere absolute Abweichung  $\delta_k$ ,  $k = 2, 3$  dieser Werte über alle Maschen des  $20 \times 20$  – Gitters ermittelt.

Die Qualität dieser Schätzer kann auch anhand der Berechnung des Korrelationskoeffizienten  $\rho$  zwischen vorhergesagten und berechneten Indizes geprüft werden, d.h. im Falle einer hohen Qualität wird folgendes erwartet:

$$\begin{aligned} \delta_{k+1} &\approx 0 \\ \rho(\hat{Y}_{k+1}, Y_{k+1}) &\approx 1 \\ k &= 1, 2 \end{aligned} \quad (8)$$

In nächstem Schritt wird eine archäologische Prognose für mögliche noch nicht entdeckte Fundorte mittels (5)–(7) sowie für  $k = 3$  durchgeführt, d.h. für das 1. Modell in 4.1

$$\hat{Y}_4 = f_{1,3} = \min(\min(B, W), Y_3) \quad (6')$$

und für das 2. Modell in 4.2

$$\hat{Y}_4 = f_{2,3} = \min(\max(B, W), Y_3) \quad (7')$$

Es soll an dieser Stelle insbesondere betont werden, dass in keinem der beiden Modelle (1) oder (2) eine Ursache-Wirkungs-Relation zwischen Variablen unterstellt wird. Deswegen ist

auch die Vorhersage mit diesen Modellen für zeitlich verschiedene Fundorte erlaubt.

## 5. Fallstudie Wetterau

### 5.1 Die Auswertung der Ergebnisse

Die Hypothese über einen Zusammenhang zwischen Variablen wird wie bereits in 4.3 beschrieben getestet. Abbildung 5 zeigt die Abhängigkeit der Testgröße  $T_k$ ,  $k = 2, 3$  vom Binarisierungsparameter  $\varepsilon$ .

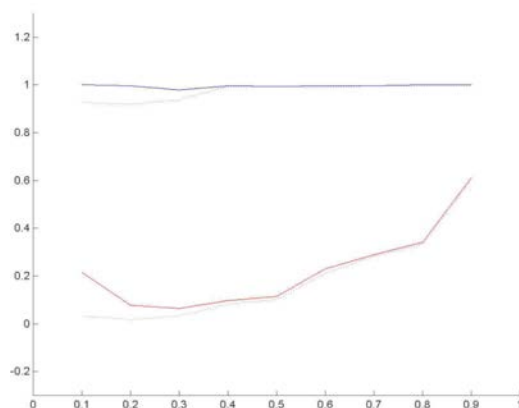


Abb. 5: Abhängigkeit der Testgröße  $T_k$ ,  $k = 2, 3$  aus (4) vom Binarisierungsparameter  $\varepsilon$ . Mit „rot“ werden die Ergebnisse für das Modell aus 4.1 und mit „blau“ die Ergebnisse für das Modell aus 4.2 dargestellt. Die durchgezogenen Linien stehen für  $k = 2$  aus (1) und (2). Die gestrichelten Linien beschreiben den Fall  $k = 1$  aus (1) und (2).

Auf dem Signifikanzniveau  $\alpha = 0.05$  kann die Hypothese über den starken Zusammenhang nach (1) für beide Epochen verworfen werden. Dagegen wird die Hypothese über den schwachen Zusammenhang nach (2) für  $\alpha = 0.1$  und beliebige Werte des Parameters  $\varepsilon$  für beide Epochen (der Urnenfelder und der römischen Kaiserzeit) nicht abgelehnt. Für die Epoche der römischen Kaiserzeit wird die Hypothese über den schwachen Zusammenhang nach (2) ebenfalls für  $\alpha = 0.05$  und beliebige Werte des Parameters  $\varepsilon$  angenommen.

Im Weiteren wird mit dem Modell aus 4.2 gearbeitet. Mittels (7) wurden Vorhersagen für die Epoche der Urnenfelder und für die Epoche der römischen Kaiserzeit erstellt. Die Vorhersage für die Epoche der römischen Kaiserzeit ist in Abbildung 6 zu sehen.

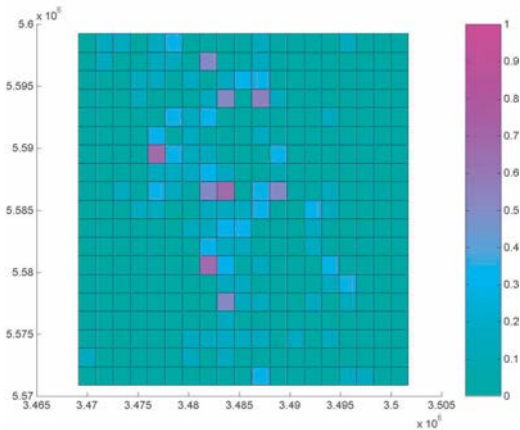


Abb. 6: Die Vorhersage  $\hat{Y}_3$  des Favorisierungsindex  $Y_3$  nach (9).

Wir erinnern daran, dass es sich dabei um 20 x 20 Werte handelt, die in (9) nochmals beschrieben sind:

$$\hat{Y}_2 = f_{2,1} = \min(\max(B, W), Y_1) \tag{9}$$

$$\hat{Y}_3 = f_{2,2} = \min(\max(B, W), Y_2)$$

Die Qualitat dieser Schatzer wird wie in (8) gepruft. Es wurde ermittelt:

$$\begin{aligned} \delta_2 &= 0.0831 \\ \delta_3 &= 0.0376 \\ \rho(\hat{Y}_2, Y_2) &= 0.11 \\ \rho(\hat{Y}_3, Y_3) &= 0.85 \end{aligned} \tag{10}$$

Die Schatzung kann vor allem fur die Epoche der romischen Kaiserzeit als akzeptabel angesehen werden. Die schwache Korrelation im Falle der Epoche der Urnenfelder spiegelt die Ungewissheit der Input-Variablen wider: Viele Siedlungen der Bandkeramiker wurden ubert und die fruheren Naturparameter konnen nur geschatzt werden.

**5.2 Eine Vorhersage mit dem LOG-Modell aus 4.2 (Wahrscheinlichkeitskarte)**

Abbildung 7 zeigt die Vorhersage (Wahrscheinlichkeitskarte) der Standorte von moglichen unentdeckten archaologischen Funde nach (7').

Mittels eines Vergleiches mit Abbildung 1 kann erkannt werden, dass eine hohe Wahrscheinlichkeit fur die Fundorte etwa den Bereichen mit einer hohen Dichte an bestehenden Fundorten entspricht. Ein Bereich hebt sich dabei deutlich von den anderen ab. Nach der Vorhersage

(Wahrscheinlichkeitskarte) der Standorte von moglichen unentdeckten archaologischen Funde kann der Bereich um den Rechtswert  $X=3.482 \times 10^6$  und den Hochwert  $Y=5.581 \times 10^6$  (Gau-Kruger-Koordinaten) fur weitere archaologische Untersuchungen empfohlen werden. Dieser Bereich entspricht auf einer topologischen Karte der naheren Umgebung von Bad Nauheim.

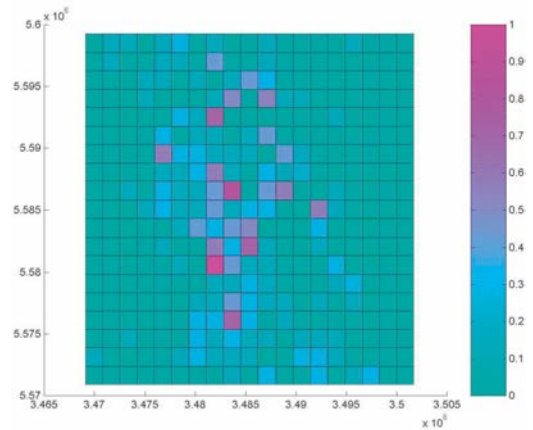


Abb. 7: Vorhersage (Wahrscheinlichkeitskarte) der Standorte von moglichen unentdeckten archaologischen Funde nach (7').

**6. Diskussion**

Wie Jackenkroll (2008) bemerkt, ist die Archaologie fur die Zeiten, aus denen keine schriftlichen Quellen existieren, auf die Arbeit in direkter Beruhrung mit den Orten angewiesen, an denen die zu erforschenden Kulturkreise lebten. Da diese zu einem Groteil im Verborgenen liegen, sind aufwendige Untersuchungen notwendig. Eine Absolutvorhersage uber das Vorkommen und Nicht-Vorkommen von Funden ist durch sie allerdings nicht moglich. Bruckner und Gerlach (2006) gehen davon aus, dass nur etwa ein Drittel der Fundstellen in Mitteleuropa bisher archaologisch gesichert ist.

Aus rein statistischer Sicht bedeutet dies, dass insbesondere keine Voraussetzungen fur die Verwendung von logischen oder anderen Regressionsansatzen erfullt sind.

Daher sollen solche Vorhersagemodelle verstarkt gesucht und weiterentwickelt werden, die wenigstens naherungsweise den Zusammenhang erfassen. Es gibt keine statistische Methode, die entweder mit komplett fehlenden oder nur vermuteten Daten effizient arbeitet. Eine Einschrankung auf moglichst ein-

fache Modelle ist empfehlenswert. Ein einfaches Modell mittels „Suitability Index“ aus Jackenkroll (2008) führte zur vergleichsweise besseren Ergebnissen als Modelle mittels logischer und linearer Regressionen. Ein Grund lag – wie bereits erwähnt – in den für diese statistischen Methoden nicht erfüllten Voraussetzungen.

Unser Vorhersagemodell ist frei von der Unterstellung einer eindeutigen Ursache-Wirkungs-Relation zwischen Variablen. Dieser Fakt erhöht die praxisbezogene Relevanz des LOG-Modells. Die Naturraumgrenzen als einer der Umweltfaktoren wurden durch den anthropogenen Eingriff in die früher dicht bewaldeten Gebiete Mitteleuropas enorm stark beeinflusst. Ebenso ist der Flussverlauf der bei Modellierung berücksichtigten Flüsse für die Zeitepoche ca. 5000 v. Chr. oft unbekannt und im Laufe der Zeit möglicherweise anthropogen verändert worden. Durch unser modifiziertes archäologisches Prognosemodell fand dieser Fakt Berücksichtigung.

Die Ergebnisse der Fallstudie zeigen eine deutliche Ablehnung der Hypothese über einen starken logisch-theoretischen Zusammenhang zwischen Variablen nach (1) und eine Annahme der Hypothese über einen schwachen Zusammenhang zwischen Variablen nach (2) auf dem Signifikanzniveau  $\alpha = 0.1$ . Speziell für die Epoche der römischen Kaiserzeit erfolgt diese Annahme bereits für  $\alpha = 0.05$ . Die Vorhersagemethode nach (5) und (7) wird anhand einer Gegenüberstellung von vorhergesagten und bestehenden Favorisierungsindizes validiert. Ihre Relevanz konnte insbesondere für die Epoche der römischen Kaiserzeit durch die in (10) präsentierten Ergebnissen nachgewiesen werden.

Nach der Vorhersage (Wahrscheinlichkeitskarte) der Standorte von möglichen unentdeckten archäologischen Funden mithilfe (7') kann der Bereich um den Rechtswert  $X = 3.482 \times 106$  und

den Hochwert  $Y = 5.581 \times 106$  (Gauß-Krüger-Koordinaten) für weitere archäologische Untersuchungen empfohlen werden, vgl. Abbildung 7. Dieser Bereich entspricht auf einer topologischen Karte der näheren Umgebung von Bad Nauheim.

#### Danksagung

An dieser Stelle möchten wir uns bei Dipl.-Geogr. Markus Jackenkroll für seine Hilfsbereitschaft bei der primären Datenaufbereitung in GIS sowie für informative Diskussionen herzlich bedanken.

#### Literatur

- [1] Brückner H. und R. Gerlach (2006): Geographie: Physische Geographie und Humangeographie. Spektrum Akademischer Verlag, Heidelberg.
- [2] Jackenkroll M. (2008): Statistische Vorhersagemodelle in einem Geographischen Informationssystem am Beispiel der Archäologie der nördlichen Wetterau. Diplomarbeit (unpubliziert). Institut für Geographie der TU Dresden.
- [3] Posluschny, A. (2002): Die hallstattzeitliche Besiedlung in Maindreieck, GIS-gestützte Fundstellenanalyse. Dissertation. Nr. 1077 in BAR International Series. Archeopress. Oxford.
- [4] Sabel K. (1982): Ursachen und Auswirkungen bodengeographischer Grenzen in der Wetterau (Hessen). Frankfurter Geowissenschaftliche Arbeiten. Bd. 3, Serie D, Fachbereich Geowissenschaften der J.W. Goethe-Universität Frankfurt. Frankfurt.
- [5] Saile T. (1998): Untersuchungen zur ur- und frühgeschichtlichen Besiedlung der nördlichen Wetterau. Materialien zur Vor- und Frühgeschichte. 21.
- [6] Van Leusen M. und H. Kamermans (Hrsg.) (2005): Predictive Modelling for Archaeological Heritage Management: A research agenda. Nederlandse Archeologische Rapporten. 29 (Amersfoort 2005).
- [7] Zimmermann A. (2002): Landschaftsarchäologie I: Die Bandkeramik auf der Aldenhovener Platte. Bericht der Rheinisch-Germanische Kommission. 83, 17-38.

#### Anschrift des Autors

PD Dr. rer. nat. habil. Olga Wälder, Institut für Kartographie, Technische Universität Dresden, Mommsenstrasse 13, D-01062 Dresden  
E-Mail: Olga.Waelder@tu-dresden.de