

## RAUM-ZEITLICHE ANALYSE VON VERBRECHENS-HOTSPOTS MITTELS TRAJEKTORIEN UND KOHONEN-KARTEN

Marco HELBICH, Julian HAGENAUER (beide Heidelberg) und  
Michael LEITNER (Baton Rouge, USA)\*

mit 4 Abb. und 2 Tab. im Text

### INHALT

<i>Summary</i> .....	291
<i>Zusammenfassung</i> .....	292
1 Hintergrund .....	292
2 Untersuchungsgebiet und Daten .....	294
3 Methodik .....	295
4 Ergebnisse .....	298
5 Schlussfolgerungen .....	302
6 Literaturverzeichnis .....	303

### *Summary*

#### *Spatio-temporal analysis of crime hotspots by trajectories and Self-Organizing Maps*

*The exploration of spatial relationships between crime incidents, the socio-economic characteristics of neighborhoods, as well as physical and structural compositions of the urban landscape is an ongoing research issue in Geographic Information Science. Spatial data mining tools improve the ability to gain knowledge from geographic data and help to understand spatio-temporal processes that contribute to the presence or absence of criminal offenses. However, most of the currently available tools focus either on the spatial, the temporal, or a combination of both aspects. But crime has a spatial and temporal component in a multidimensional attribute space. Therefore, it is reasonable to combine all these aspects within one analytical framework.*

---

\* Mag. Dr. Marco HELBICH, Dipl.-Inf. Julian HAGENAUER, Geographisches Institut, Universität Heidelberg, Berliner Straße 48, D-69120 Heidelberg, Deutschland; E-Mail: helbich@uni-heidelberg.de, hagenauer@uni-heidelberg.de, <http://giscience.uni-hd.de>; Assoc.-Prof. Dr. Michael LEITNER, Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA; E-Mail: mleitne@lsu.edu, <http://www.ga.lsu.edu/leitner.html>; Mag. Dr. Marco HELBICH wurde von der Alexander-von-Humboldt-Stiftung finanziert.

*This paper presents such a methodology to explore crime patterns and their spatial and temporal behavior within their socio-economic and environmental neighborhoods. The framework consists of three complementary techniques: A spatio-temporal scan statistic to detect crime hotspots, a growing Self-Organizing Map (SOM) to analyze attribute properties of the neighborhoods, and mapping of crime hotspot trajectories onto different SOM visualizations. The case study uses burglary locations from Houston, Texas, from August to October 2005.*

## Zusammenfassung

*Die Exploration räumlicher Beziehungen zwischen Kriminalität, sozio-ökonomischen und raumstrukturellen Merkmalen administrativer Einheiten ist ein aufstrebendes Forschungsfeld der Geoinformation. Räumliches Data Mining trägt dabei maßgeblich zum besseren Verständnis von geographischen Daten und raum-zeitlichen Prozessen bei. Die meisten Verfahren konzentrieren sich entweder auf räumliche Aspekte, zeitliche Aspekte oder eine Kombination von beiden, wobei sich Straftaten aber in einem mehrdimensionalen Merkmalsraum ereignen. Deshalb ist es notwendig, diese multidimensionalen Aspekte zu kombinieren und simultan zu betrachten.*

*Dieser Beitrag präsentiert erstmals einen solchen Ansatz zur Exploration von raum-zeitlichen Kriminalitätsmustern unter Berücksichtigung des sozio-ökonomischen und demographischen Kontextes. Das Analysedesign besteht aus der raum-zeitlichen Scan-Statistik zur Detektion von Kriminalitäts-Hotspots, einer wachsenden KOHONEN-Karte (Self-Organizing Map, SOM) zur Klassifikation des Kontextes und Trajektorien zur Visualisierung von zeitlichen Veränderungen der Verbrechens-Hotspots. Das Fallbeispiel analysiert Einbrüche während der Hurrikansaison in der Stadtregion Houston, Texas, im Zeitraum August bis Oktober 2005.*

## 1 Hintergrund

In unserem Informationszeitalter bedarf es effizienter numerischer Algorithmen und visueller analytischer Zugänge, um die großen Mengen an Geodaten adäquat analysieren und visualisieren zu können (KEIM et al. 2008, MILLER & HAN 2009, ANDRIENKO et al. 2010). Dies gilt besonders auch für kriminologische Analysen. Heute sind klassische Visualisierungsansätze, wie sie zum Beispiel ECK et al. (2005) oder CHAINEY & RATCLIFFE (2005) diskutieren, nur noch bedingt anwendbar, da diese Analyseverfahren zumeist ausschließlich die räumliche Komponente betrachten (z.B. Geographical Analysis Machine [GAM] von OPENSHAW et al. 1987). Kerndichteschätzungen (GATRELL 1994) etwa sind ein gängiges Verfahren zur Analyse von Verbrechens-Hotspots (z.B. LEITNER & HELBICH 2011). Diese Technik berücksichtigt jedoch ausschließlich den räumlichen Kontext von Kriminaldelikten. Die zeitliche und attributive Komponente bleiben gänzlich unberücksichtigt. Die mehrfache Dimensionalität aktueller Datensätze bedarf moderner Methoden, um Raum, Zeit und Attributinformatoren effizient verarbeiten

zu können. Data-Mining-Methoden erweitern in diesem Zusammenhang das Methodenspektrum der Kriminalanalytik (z.B. HELBICH & LEITNER 2011) und erlauben es, Rückschlüsse oder Wissen aus raum-zeitlichen kriminologischen Prozessen abzuleiten.

Bisher hat sich nur eine begrenzte Anzahl von Arbeiten mit derartigen Ansätzen befasst. Beispiele sind die Anwendung von Geovisualisierungsansätzen (z.B. Isoflächen) zur Darstellung raum-zeitlicher Kriminalitätsmuster. Die Evaluierung ihrer Tauglichkeit ist Bestandteil heutiger Forschung (BRUNSDON et al. 2007). Zur web-basierten Analyse von raum-zeitlichen Kriminalitätsverteilungen wurde von ROTH et al. (2010) das Visualisierungstool Geovista CrimeViz entwickelt. NAKAYA & YANO (2010) präsentieren einen dreidimensionalen Kartierungsansatz innerhalb des Raum-Zeit-Würfels. Die Autoren kombinieren dabei Kerndichteschätzung (GATRELL 1994) und Scan-Statistik (KULLDORF 1997). Dies ermöglicht eine simultane Abbildung der räumlichen Ausdehnung sowie die Analyse der Dauer von Kriminalitäts-Hotspots.

In einer komparativen Analyse unterschiedlicher Methoden zur Clusterdetektion belegen HELBICH & LEITNER (2011), dass insbesondere die Scan-Statistik im Vergleich zur GAM und der Statistik von BESAG & NEWELL (1991) zur Erforschung von Kriminalitäts-Hotspots geeignet ist. ANDRIENKO et al. (2010) propagieren einen integrativen Ansatz bestehend aus KOHONEN-Karten (Self-Organizing Maps, SOMs, KOHONEN 1982, 2001), gekoppelt mit einer Reihe von interaktiven Visualisierungstools. Bei sämtlichen genannten Arbeiten liegt der Schwerpunkt entweder auf der räumlichen Komponente, der zeitlichen Komponente oder einer Kombination von Raum und Zeit. Kriminalitäts-Hotspots unterliegen aber auch umwelt- und kontextbedingten Einflüssen. Die jeweiligen sozio-ökonomischen Eigenschaften eines Gebiets hemmen oder fördern das Auftreten von Delikten. Es ist daher unabdingbar, die Aspekte „Raum“, „Zeit“ und „Kontext“ in einen methodischen Rahmen zu fassen und simultan zu analysieren.

Vor diesem Hintergrund soll mit diesem Beitrag ein alternativer methodischer Ansatz zur Verbrechensanalyse vorgestellt werden, mit dem räumliche und zeitliche Eigenschaften von Verbrechen innerhalb ihres sozio-ökonomischen Kontextes analysiert werden können. Der Ansatz vereint erstmals folgende komplementäre Techniken: (a) eine raum-zeitliche Scan-Statistik zur Detektion von Kriminalitäts-Hotspots; (b) eine Klassifikation des Kontextes auf Basis einer wachsenden SOM und (c) die gekoppelte Visualisierung der Hotspots auf der SOM mittels Trajektorien. Als Anwendungsbeispiel dient das Verteilungsmuster von Einbruchdelikten innerhalb der Stadtregion von Houston, Texas, während der Hurrikansaison 2005.

Aktuelle empirische Ergebnisse (z.B. LAWTON & SCHULENBURG 2007, LEITNER & HELBICH 2011) belegen den direkten Zusammenhang zwischen Hurrikans und der räumlichen Verteilung von Kriminalitätsdelikten. So sinkt im Allgemeinen die Kriminalität nach dem Eintreffen eines Hurrikans, um später wieder auf das Kriminalitätsniveau von vor dem Eintreffen des Hurrikans anzusteigen. Dies gilt für die vom Hurrikan direkt betroffenen Regionen. Im Gegensatz dazu stagniert oder sinkt die Kriminalität leicht in Regionen, die aus den von der Naturkatastrophe betroffenen Gebieten evakuierte Personen aufgenommen haben (z.B. LEITNER & HELBICH 2011, LEITNER et al. 2011).

Für die Stadt Houston haben frühere Analysen gezeigt, dass Hurrikan Rita im September 2005 im Gegensatz zu Hurrikan Katrina im August 2005 zu einem kurzfristigen

Anstieg von Einbrüchen führte. Konträr dazu nahm die Anzahl anderer Deliktarten (Raubüberfälle, Morde etc.) leicht ab (LEITNER & HELBICH 2009, 2011). Die Autoren führen diesen Anstieg auf die behördlich angeordneten Evakuierungsmaßnahmen vor Hurrikan Rita zurück. Bei Hurrikan Katrina blieb hingegen eine Evakuierung aus.

Aus räumlicher Perspektive kam es während des Hurrikans Rita vor allem in den östlichen Teilen der Metropolregion Houston, die einen hohen Anteil an afroamerikanischer und lateinamerikanischer Bevölkerung beherbergen, zu einem rapiden und massiven Anstieg an Einbrüchen. LEITNER & HELBICH (2011) erklärten dies damit, dass Personen, die nicht der Evakuierung folgten, die Einbrüche in den nahe gelegenen und leerstehenden Gebäuden begangen haben. Die Arbeit von LEITNER & HELBICH (2011) beschränkt sich jedoch ausschließlich auf die Analyse raum-zeitlicher Verteilungsmuster, ein expliziter Einbezug des multivariaten demographischen und sozio-ökonomischen Kontextes fehlt. Der vorliegende Artikel schließt diese Forschungslücke.

Im Detail sollen im Zuge dieser Analyse folgende Fragen geklärt werden: Ist die vorgeschlagene Methodik zur Analyse und Kartierung von Straftaten geeignet? Welche Gemeinsamkeiten und Unterschiede der Verbrechens-Hotspots lassen sich während der Hurrikans Katrina und Rita feststellen? Welche Auswirkungen haben sozio-ökonomische und demographische Unterschiede von administrativen Einheiten (in diesem Fall: Nachbarschaften) auf Verbrechens-Hotspots? Lassen sich bei den Trajektorien zyklische Muster oder plötzliche Veränderungen feststellen?

Dieser Artikel ist wie folgt strukturiert: In Abschnitt 2 werden das Untersuchungsgebiet und der zugrunde liegende Datensatz vorgestellt. Abschnitt 3 skizziert die angewandte Methodik. Abschnitt 4 diskutiert die resultierenden Modellergebnisse. Eine abschließende Einordnung der Ergebnisse in den theoretischen Kontext erfolgt in Abschnitt 5.

## 2 Untersuchungsgebiet und Daten

Die Metropolregion Houston, Texas, bildet das Untersuchungsgebiet. Sie wurde immer wieder von Hurrikans betroffen, zuletzt durch die beiden Hurrikans Katrina und Rita im Jahr 2005. Um deren Auswirkung auf die Kriminalität zu analysieren, wurden Verbrechensdaten vom Houston Police Department herangezogen. Der Untersuchungszeitraum umfasst wenige Wochen vor, während und nachdem die Hurrikans Katrina (29. August 2005) und Rita (24. September 2005) das Festland in der Nähe von Houston erreichten. Im Anschluss an die Geokodierung mittels des TIGER-Straßengraphens (US Census Bureau) standen Daten von rund 25.000 Straftaten (Einbruch, Autodiebstahl, Morde, schwere Körperverletzung etc.) zur Verfügung. Eine zeitliche Analyse, differenziert nach Deliktarten, zeigt eine hohe Kontinuität der Delikte (LEITNER & HELBICH 2009, 2011). Eine Ausnahme bilden Einbrüche, deren Zahl während des Hurrikans Rita signifikant anstieg. Aus diesem Grund konzentriert sich die vorliegende Analyse ausschließlich auf die raum-zeitliche Verteilung der insgesamt 6.730 Einbruchsdelikte.

Ergänzend werden sozio-ökonomische, demographische und raumstrukturelle Variablen zur näheren Charakterisierung des Kontextes und der Nachbarschaften

[neighborhoods] herangezogen, da davon auszugehen ist, dass Houston von wirtschaftlichen und ethnischen Segregationstendenzen geprägt ist, die im Zusammenhang mit der Einbruchsverteilung stehen. Vorrangegangene Regressionsanalysen zeigten, dass insbesondere die in Tabelle 1 angeführten Variablen als maßgebend für die Einbruchsdichte angesehen werden können (LEITNER & HELBICH 2011). Deshalb werden zur Charakterisierung der Nachbarschaften ausschließlich diese vier Variablen auf Basis von 408 administrativen Einheiten [census tracts] verwendet.

Variablen	Min.	1. Quantil	Median	Mittelwert	3. Quantil	Max.
Anteil afroamerikanischer Bevölkerung an der Gesamtbevölkerung im Jahr 2000 in %	0,00	2,83	10,57	25,56	38,72	98,26
Anteil lateinamerikanischer Bevölkerung an der Gesamtbevölkerung im Jahr 2000 in %	0,00	10,89	27,91	35,05	54,16	97,16
Anteil von Personen unter der Armutsgrenze an der Gesamtbevölkerung im Jahr 1999 in %	0,00	8,15	18,30	18,84	26,55	70,40
Euklidische Distanz zur nächsten Polizeidienststelle (in Metern; Stand 2009)	99	1.563	2.343	2.739	3.564	9.822

Quelle: US Census Bureau

Tab. 1: Sozio-ökonomische, demographische und raumstrukturelle Variablen zur Charakterisierung der administrativen Einheiten

### 3 Methodik

Abbildung 1 fasst die wesentlichen Komponenten des Analysedesigns zusammen. Der erste Schritt zielt mittels Scan-Statistik darauf ab, raum-zeitliche Kriminalitäts-Hotspots zu identifizieren (detektieren). Im zweiten Schritt wird anhand einer SOM der räumliche Kontext klassifiziert. Abschließend werden die raum-zeitlichen Hotspots als Trajektorien auf SOM-Visualisierungsformen dargestellt.

#### 3.1 Raum-zeitliche Scan-Statistik

Die Scan-Statistik (KULLDORFF 1997, KULLDORFF et al. 1998) eignet sich zur Detektion von raum-zeitlichen Hotspots. In der Kriminalitätsanalyse wurde sie bisher aber nur selten angewendet (z.B. LEITNER & HELBICH 2011). Bei einem raum-zeitlichen Hotspot handelt es sich um eine nicht zufällige, räumlich und zeitlich begrenzte Ballung von Delikten (in Anlehnung an KNOX 1989).

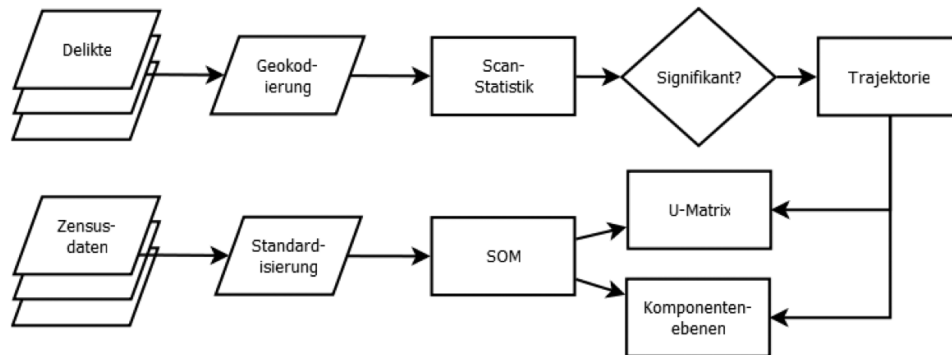


Abb. 1: Analysedesign

Der Scan-Statistik dienen die Kriminalitätsstandorte und deren Zeitstempel als Eingangsdaten. Ein als Zylinder definiertes Scan-Fenster bewegt sich von Delikt  $i$  zu Delikt  $j$ , wobei die Grundfläche des Zylinders die räumliche Dimension und dessen Höhe die zeitliche Dimension repräsentieren. Dabei verändert das Scan-Fenster kontinuierlich seinen Radius und seine Höhe. Dies resultiert in einer hohen Anzahl unterschiedlich großer und sich überschneidender Zylinder, bei denen die beobachtete Zahl der Einbrüche mit der erwarteten Zahl auf Basis eines raum-zeitlichen Permutationsmodells verglichen wird (KULLDORFF et al. 1998). Die zu erwartende Anzahl an Straftaten wird mittels Poisson-Verteilung geschätzt. Ein raum-zeitlicher Hotspot ist somit ein Zylinder, der eine signifikant höhere Zahl an beobachteten Fällen aufweist als man zufällig für einen bestimmten Zeitraum erwarten könnte. Mithilfe von Monte-Carlo-Simulationen wird die statistische Signifikanz der einzelnen Hotspots ermittelt.

### 3.2 SOM-Algorithmus und Growing-Grid-Erweiterung

SOMs sind ein von KOHONEN (1982, 2001) entwickeltes unüberwachtes Lernverfahren auf Grundlage eines neuronalen Netzes. Es erlaubt die nicht-lineare Abbildung eines hochdimensionalen Eingaberaumes auf einem zumeist zwei-dimensionalen Ausgaberaum. Im Zuge des iterativen Algorithmus bleibt die Topologie des Eingaberaumes weitgehend erhalten. Inputvektoren, die im Eingaberaum Ähnlichkeit aufweisen, werden auch im Ausgaberaum in gewisser „räumlicher“ Nähe abgebildet. SOMs eignen sich zur Visualisierung (VESANTO 1999), zur Vektorquantifizierung und zum Clustering (VESANTO & ALHONIEMI 2000). In der Geographie sind SOMs in erster Linie zur Klassifikation von räumlichen Prozessen eingesetzt worden (z.B. AGARWAL & SKUPIN 2008, SPIELMAN & THILL 2008, ARRIBAS-BEL et al. 2011).

SOMs bestehen aus einer beliebigen Anzahl von Neuronen, die mit benachbarten Neuronen verbunden sind. Diese Beziehungen definieren die Topologie der Karte.

Jedes Neuron wird durch einen Prototypvektor, der die gleiche Dimension wie der Eingaberaum hat, repräsentiert. Bei jedem Lernschritt wird nun ein zufällig gewähltes Sample mit der Menge an Inputvektoren, die zufällig initialisiert sind, verglichen. Jenes Neuron, das den geringsten euklidischen Abstand zum Eingabevektor aufweist, wird als Gewinnerneuron (Best Matching Unit, BMU) deklariert. Nun werden bei der BMU sowie in reduzierter Form auch bei ihrem topologischen Nachbarn, die Gewichte adaptiert. Dies hat zur Folge, dass sich nach etlichen Iterationen die BMUs und deren Nachbarn auf den Eingaberaum zu bewegen und so den ursprünglichen Datenraum abbilden. Im Zuge der Trainingsphase wird die Lernrate kontinuierlich gesenkt. Der Lernprozess geht dabei schrittweise von einer ersten groben Lernphase, in der die Neuronen nur grob auf der Karte angeordnet werden, zu einer Phase der Feinabstimmung mit geringfügigen Änderungen über. Ausführliche methodische Details sind KOHONEN (2001) zu entnehmen. Nachteile des klassischen SOM-Algorithmus, der zumeist im GIS-Kontext verwendet wird (z.B. SKUPIN & HAGELMAN 2005, SPIELMAN & THILL 2008, ARRIBAS-BEL et al. 2011), sind die subjektive Wahl der Modellparameter und die Spezifikation der Kartendimension.

FRITZKE (1995) schlug deshalb ein wachsendes Netz (Growing Grid, GG) mit dynamischer Kartendimension vor. Normalerweise beginnt das Training eines GGs mit einer sehr kleinen (zweidimensionalen) Karte. Während des Trainings wird jenes Neuron, das die meisten Eingabedaten abbildet, um eine Zeile oder Spalte ergänzt. Dabei wird zwischen diesem Neuron und dem in dessen Nachbarschaft unähnlichsten Neuron eine neue Spalte/Zeile an Neuronen eingefügt und so die Dichte der Prototypvektoren in diesem Bereich erhöht. Dies hat zur Folge, dass der Eingangsraum gleichmäßiger repräsentiert wird. Das Wachsen des GGs wird so lange wiederholt, bis ein Abbruchkriterium erreicht wird (z.B. wenn die Karte zu einer maximalen Größe gewachsen ist).

### 3.3 Visualisierung raum-zeitlicher Veränderungen mittels Trajektorien

Eine SOM erlaubt eine statische Zuordnung des Eingangsraumes auf einem 2D-Raster, um so Informationen über den Eingaberaum (z.B. Cluster in den Daten) zu gewinnen (VESANTO & ALHONIEMI 2000). Die meisten Daten haben sowohl eine räumliche als auch eine zeitliche Komponente. Zweitens wird bei Visualisierungen aber oftmals vernachlässigt.

Eine pragmatische Lösung besteht darin, eine separate SOM-Visualisierung für jeden Zeitstempel zu berechnen, um so auf zeitliche Veränderungen schließen zu können. ANDRIENKO et al. (2010) koppeln Visual Analytics (KEIM et al. 2008) mit SOMs zur Analyse von raum-zeitlichen Daten. Ein alternativer Ansatz zur Visualisierung von zeitlichen Mustern ist die Verwendung von Trajektorien in Kombination mit SOMs (SKUPIN & HAGELMAN 2005). Eine Trajektorie repräsentiert dabei den Verlauf eines Objektes auf einer SOM und erlaubt Rückschlüsse auf dessen zeitliche Entwicklung in Abhängigkeit vom attributiven Kontext.

## 4 Ergebnisse

### 4.1 Detektion von Verbrechens-Hotspots

Die durchschnittliche Zahl der Einbrüche pro Woche im Zeitraum August bis Oktober 2005 lag in der Stadt Houston bei ca. 450 Fällen und zeigt mit einer Ausnahme wenig Fluktuation. Einen dramatischen Anstieg gab es in der Woche zwischen dem 18. und 24. September. In dieser Woche verdreifachte sich die Zahl der Einbrüche nahezu auf 1.185 Delikte.

Die Scan-Statistik wurde zur Identifikation von raum-zeitlichen Einbruchs-Hotspots herangezogen. Für den gesamten Zeitraum gab es mehrere raum-zeitliche Hotspots. Vier dieser Hotspots (vgl. Tab. 2) weisen ein Signifikanzniveau von  $p < 0,05$  auf. Das Signifikanzniveau basiert auf 999 Monte-Carlo-Simulationsläufen. Der Großteil der Hotspots existierte lediglich einen Tag. Hurrikan Katrina, der Houston am 29. August 2005 traf, hatte keine signifikanten Auswirkungen auf die raum-zeitliche Verteilung der Einbruchsdelikte. Keiner der detektierten Hotspots entspricht dem Zeitpunkt des Auftretens von Hurrikan Katrina. Konträr dazu verhielt es sich beim Hurrikan Rita, wo im Zeitraum zwischen dem 21. bis 27. September 2005 ein hoch signifikanter Hotspot ( $p < 0,001$ ) festgestellt werden konnte. Es führte also die Zwangsevakuierung der Bevölkerung zu einem signifikanten Anstieg an Einbruchsdelikten. Im Vergleich zu den anderen Hotspots weist dieser auch eine beachtliche räumliche Ausdehnung auf und umfasst insgesamt 357 Einbrüche. Eine Zusammenfassung der vier statistisch signifikanten Hotspots mit  $p < 0,05$  ist Tabelle 2 zu entnehmen.

Die Diskussion beschränkt sich in der Folge auf diese vier signifikanten Hotspots. In einem ersten Schritt wurde das räumliche Zentrum (Durchschnitt der  $xy$ -Koordinaten der Einbrüche in Hotspot  $i$ ) der Hotspots berechnet. Anschließend wurden jene Census Tracts ermittelt, in welche die vier Hotspotzentren fallen und im Anschluss als Trajektorie auf der SOM visualisiert. Abbildung 2 zeigt die Zentren der vier Hotspots sowie deren sequenzielles Auftreten. Der erste Hotspot tritt im Westen des Stadtgebiets in Erscheinung, und die beiden nachfolgenden Hotspots weiter ostwärts. Der vierte Hotspot befindet sich wieder in unmittelbarer Nähe zum zweiten Hotspot, im Zentrum von Houston.

	Anfang des Hotspots	Ende des Hotspots	Anzahl der Delikte	$p$ -Wert
1	25.08.2005	25.08.2005	4	0,048
2	09.09.2005	09.09.2005	6	0,019
3	21.09.2005	27.09.2005	357	0,001
4	10.10.2005	10.10.2005	5	0,008

Tab. 2: Ergebnisse der raum-zeitlichen Scan-Statistik





Abb. 2: Mittelpunkte der Einbruchs-Hotspots in zeitlicher Ordnung analog zu Tabelle 2 (Hintergrundkarte OpenStreetMap)

#### 4.2 Klassifikation des Kontextes und Visualisierung der raum-zeitlichen Dynamik der Hotspots

Um das Auftreten von Hotspots im Attributraum zu analysieren und den demographischen und sozio-ökonomischen Kontext zu berücksichtigen, wurde ein Growing Grid (GG) berechnet. Wie in Abschnitt 2 diskutiert, kam es zu einer Vorauswahl von Attributen (vgl. Tab. 1). Das GG konnte so lange wachsen, bis 500 Neuronen überschritten wurden. Die daraus resultierende Karte besteht aus einem Raster mit 24 Spalten und 21 Zeilen. Da in der Interpretation zwischen GG und SOM kein Unterschied besteht, wird in weiterer Folge auf eine Differenzierung verzichtet.

Ein Vorteil der SOM ist ihre flexible Visualisierung (VESANTO 1999). Im Folgenden werden zwei Methoden betrachtet: (a) die U-Matrix (Unified Distance Matrix; ULTSCH 1993) und (b) die Komponenten-Ebenen (Component Planes, CPs). Die U-Matrix visualisiert die Differenz benachbarter Neuronen auf der Karte. Bedingt durch die Anzahl der Nachbarneuronen ist hierfür eine größere Zahl an Elementen notwendig als die SOM Neuronen besitzt. Die U-Matrix ist geeignet, Cluster und deren Qualität in den Daten zu charakterisieren. Eine CP ist eine Projektion einer einzigen Komponente (bzw. eines Nachbarschaftsattributes) auf die SOM. Alle Komponenten bilden zusammen die gesamte Information der Karte. Durch die visuelle Analyse einzelner CPs oder im Vergleich mit anderen CPs der SOM können Informationen über die SOM-Struktur gewonnen werden.

Abbildung 3 zeigt die U-Matrix der trainierten SOM, wobei keine klaren Cluster-grenzen zu erkennen sind. Dennoch erscheint am rechten Rand der U-Matrix eine homogenere Fläche. Deren genauere Analyse ergibt, dass dunkelgraue Zellen geringe Anteile afroamerikanischer und lateinamerikanischer Bevölkerung und einen geringen Anteil an Personen unter der Armutsgrenze haben.

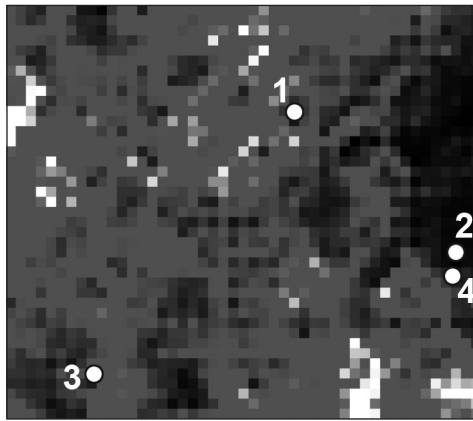


Abb. 3: U-Matrix und Trajektorie der Hotspots. Weiße Zellen haben hohe, schwarze Zellen geringe Distanz zwischen den Neuronen

Ein Vergleich zwischen den CPs (vgl. Abb. 4) zeigt, dass Armut in erster Linie mit dem Anteil an afroamerikanischer Bevölkerung und geringfügig mit dem Anteil an lateinamerikanischer Bevölkerung korreliert. Die CP der Variable „Entfernung bis zur nächsten Polizeidienststelle“ ist mit den restlichen CPs nicht korreliert.

Basierend auf den Ergebnissen der Scan-Statistik und der SOM kann eine Trajektorie auf die SOM bzw. deren U-Matrix sowie auf die einzelnen CPs projiziert werden. Da die eruierten Hotspots unterschiedliche zeitlich-diskrete Phänomene repräsentieren und sich nicht kontinuierlich über Raum und Zeit erstrecken, wird deren Trajektorie im Folgenden – konträr zu SKUPIN & HAGELMAN (2005) – nicht als Pfad, sondern durch Punkte visualisiert (vgl. Abb. 3 und 4). Die Nummerierungen entsprechen dabei dem sequenziellen Auftreten der Hotspots (vgl. Tab. 2).

Die Trajektorie beginnt am 25. August 2005, vier Tage vor dem Eintreffen von Hurrikan Katrina in Houston (29. August), und endet am 10. Oktober, mehrere Wochen nach dem Eintreffen von Hurrikan Rita (24. September). Hurrikan Rita ereignete sich während des dritten Hotspots. Die Trajektorie zeigt dabei ein markantes Muster. Einerseits befindet sie sich beim zweiten und vierten Hotspot an nahezu identer Stelle, wobei diese beiden Hotspots auch im geographischen Raum benachbart sind (vgl. Abb. 2). Andererseits sind der erste und letzte Hotspot deutlich voneinander getrennt, was besagt, dass sie in unterschiedlichen Attributräumen auftraten. Interessant ist, dass der zweite und vierte Hotspot ähnliche Attribute aufweisen, was daran erkennbar ist, dass sie auf der SOM räumlich benachbart sind. Dagegen sind sich der erste und letzte Hotspot hinsichtlich ihrer Attribute (vgl. Tab. 1) sehr unähnlich, sodass diese beiden Hotspots in unterschiedlichen Bereichen der SOM zu liegen kommen.

Kombiniert man die Trajektorie mit der U-Matrix (vgl. Abb. 3), so zeigt sich, dass die Hotspots 2 und 4 in Gebieten mit geringer Armutsrate und geringen Prozentsätzen afro- sowie lateinamerikanischer Bevölkerung liegen. Die Hotspots 1 und 3 sind in heterogenen Gebieten lokalisiert, was eine eindeutige Charakterisierung des Kontextes erschwert. Visualisiert man die Trajektorie auf den CPs (vgl. Abb. 4), zeigt sich für den Anteil an afroamerikanischer Bevölkerung kein signifikantes Muster. Die Hot-

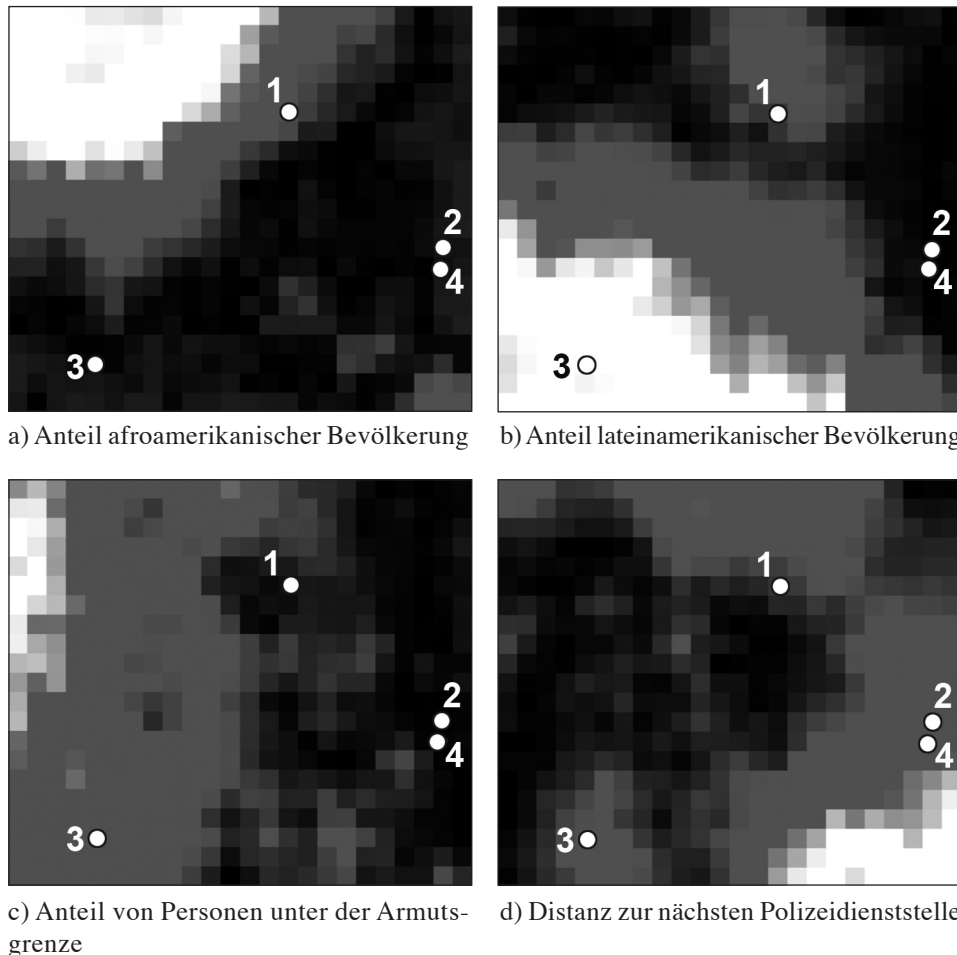


Abb. 4: CPs und Trajektorie der Hotspots. Weiße Zellen zeigen hohe, schwarze Zellen geringe Werte der einzelnen Attribute

spots 2, 3 und 4 befinden sich in Gebieten mit geringem Anteil afroamerikanischer Bevölkerung. Lediglich der erste Hotspot befindet sich innerhalb von Census Tracts mit höheren Prozentanteilen afroamerikanischer Bevölkerung.

Beim Anteil lateinamerikanischer Bevölkerung weist die Trajektorie ein signifikantes Muster auf. Während der zweite und vierte Hotspot in Regionen mit geringem Anteil lateinamerikanischer Bevölkerung liegen, ist der dritte Hotspot in einem Gebiet mit hohem Anteil gelegen. Da das zeitliche Auftreten des Hotspots mit dem Eintreffen des Hurrikans Rita in Houston übereinstimmt, lässt dies die Vermutung zu, dass die Zwangsevakuierung einen maßgeblichen Einfluss auf die raum-zeitliche Verteilung hatte. Diese Vermutung wird von den Ergebnissen von LEITNER & HELBICH (2011) gestützt, die in diesem Zeitraum einen erheblichen Anstieg an Einbruchsdelikten in lateinamerikanisch geprägten Nachbarschaften festgestellt haben. Ein hoher Anteil

an Personen unterhalb der Armutsgrenze scheint für Gebiete mit einem hohen Anteil afro- und lateinamerikanischer Bevölkerung charakteristisch zu sein. Deshalb ist es auch nicht verwunderlich, dass der Hotspot während des Hurrikans Rita diese Regionen prägt. Im Gegensatz dazu treten die anderen drei Hotspots in Gebieten mit einem niederen Prozentsatz an Personen unterhalb der Armutsgrenze auf. Keinen Einfluss hat hingegen die Entfernung zur nächsten Polizeistation.

## 5 Schlussfolgerungen

In diesem Beitrag wurde ein raum-zeitliches Analysedesign, bestehend aus Scan-Statistik, SOM und Trajektorien, zur Analyse von Kriminalitäts-Hotspots vorgestellt. Dabei wurden die zeitliche und räumliche Dimension von Hotspots, aber auch deren multivariater sozio-ökonomischer und demographischer Kontext explizit berücksichtigt. Jede dieser Komponenten bietet für sich die Möglichkeit, einzelne Aspekte von Kriminalitätsverteilungen zu analysieren: Die Scan-Statistik ermöglicht es, räumlich-zeitliche Hotspots zu detektieren, während die SOM und Trajektorien für die Analyse zeitlicher multivariater Muster hilfreich sind. Die Kombination dieser Techniken induziert Synergien und ermöglicht durch das simultane Erforschen von Raum, Zeit und Kontext eine noch leistungsfähigere Analyse, die zum besseren Verständnis räumlicher und zeitlicher Prozesse beiträgt.

Die Nützlichkeit dieses Ansatzes wurde am Fallbeispiel von Einbruchsdelikten in Houston demonstriert. Es zeigte sich, dass sich der Hotspot an Einbruchsdelikten während des Hurrikans Rita signifikant von vorigen und nachfolgenden Hotspots unterscheidet. Dies gilt für dessen räumliche Ausdehnung, aber auch für die Charakteristika der betroffenen Gebiete. Im Gegensatz zu Hurrikan Rita konnte bei Hurrikan Katrina kein Hotspot ermittelt werden. Die Tatsache, dass bei Hurrikan Rita eine Zwangsevakuiierung (im Gegensatz zu Hurrikan Katrina) angeordnet wurde, lässt den Schluss zu, dass diese maßgeblich die Anzahl und Verteilung der Einbruchsdelikte beeinflusst hat. Die Analyse der SOM und deren Trajektorie ergaben ferner, dass der Hotspot während des Hurrikans Rita in Gebieten mit differenter ethnischer Zusammensetzung sowie von Armut geprägten Gebieten auftrat. Diese Schlussfolgerungen bestätigen die Ergebnisse von LEITNER & HELBICH (2011) und verdeutlichen die Komplementarität von explorativen und inferenzstatistischen Methoden.

Der Analyseansatz ist keinesfalls auf das herangezogene Untersuchungsgebiet limitiert. Zur Übertragbarkeit auf beliebige Untersuchungsgebiete (und Disziplinen) benötigt man einerseits geokodierte Punktdaten inklusive Zeitstempel und andererseits Daten zur Charakterisierung der administrativen Einheiten (z.B. Daten der amtlichen Statistik). Zukünftige Arbeiten müssten systematisch weitere Deliktarten analysieren, um Gemeinsamkeiten oder Divergenzen aufzuzeigen. Dazu bedarf es weiterer sozio-ökonomischer und demographischer Variablen, um zu einem detaillierteren Verständnis hinsichtlich des Zusammenhangs von Hurrikans und Kriminalität zu gelangen. Ebenso muss die Effizienz der wachsenden SOM im Vergleich zum Basisalgorithmus oder zu hierarchischen SOMs evaluiert werden.

## 6 Literaturverzeichnis

- AGARWAL P., SKUPIN A. (2008), *Self-Organising Maps: Applications in Geographic Information Science: Applications in GIScience*. West Sussex, Wiley.
- ANDRIENKO G., ANDRIENKO N., BREMM S., SCHRECK T., VON LANDESBERGER T., KEIM D.A., BAK P. (2010), Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatio-temporal Patterns. In: *EuroVis*, S. 913–922.
- ARRIBAS-BEL D., NIJKAMP P., SCHOLTEN H. (2011), Multidimensional Urban Sprawl in Europe: A Self-Organizing Map Approach. In: *Computers, Environment and Urban Systems*, 35, S. 263–275.
- BESAG J., NEWELL J. (1991), The Detection of Clusters in Rare Diseases. In: *Journal of the Royal Statistical Society A*, 154, S. 143–155.
- BRUNSDON C., CORCORAN J., HIGGS G. (2007), Visualising Space and Time in Crime Patterns: A Comparison of Methods. In: *Computers, Environment and Urban Systems*, 31, 1, S. 52–75.
- CHAINEDY S., RATCLIFFE J.H. (2005), *GIS and Crime Mapping*. Chichester, Wiley.
- ECK J., CHAINEDY S., CAMERON J., LEITNER M., WILSON R. (2005), *Mapping Crime: Understanding Hotspots*. Washington DC, National Inst. of Justice.
- FRTZKE B. (1995), Growing Grid – A Self-Organizing Network with Constant Neighborhood Range and Adaptation Strength. In: *Neural Processing Letters*, 2, 5, S. 9–13.
- GATRELL A. (1994), Density Estimation and the Visualization of Point Patterns. In: HEARNshaw H., UNWIN D. (Hrsg.), *Visualization in Geographical Information Systems*, S. 65–75. Chichester, Wiley.
- HELBICH M., LEITNER M. (2011), Evaluation of Spatial Cluster Detection Algorithms for Crime Locations. In: GAUL W., GEYER-SCHULZ A., SCHMIDT-THIEME L., KUNZE J. (Hrsg.), *Challenges at the Interface of Data Analysis, Computer Science, and Optimization. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin, Springer (im Erscheinen).
- KEIM D.A., MANSMANN F., SCHNEIDEWIND J., ZIEGLER H., THOMAS J. (2008), Visual Analytics: Scope and Challenges. In: SIMOFF S.J., BÖHLEN M.H., MAZEIKA A. (Hrsg.), *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics (= Lecture Notes in Computer Science)*, S. 76–90. Heidelberg, Springer.
- KNOX E. (1989), Detection of Clusters. In: ELLIOTT P. (Hrsg.), *Methodology of Enquiries into Disease Clustering*. Small Area Health Statistics Unit, London, S. 17–20.
- KOHONEN T. (1982), Self-organized Formation of Topologically Correct Feature Maps. In: *Biological Cybernetics*, 43, S. 59–69.
- KOHONEN T. (2001), *Self-Organizing Maps*. New York, Springer.
- KULLDORF M. (1997), A Spatial Scan Statistic. In: *Communications in Statistics: Theory and Methods*, 26, S. 1481–1496.
- KULLDORFF M., ATHAS W., FEUER E., MILLER B., KEY C. (1998), Evaluating Cluster Alarms: A Space-Time Scan Statistic and Brain Cancer in Los Alamos. In: *American Journal of Public Health*, 88, S. 1377–1380.
- LAWTON B.A., SCHULENBURG J. (2007), Assessing the Impact of Hurricane Katrina on space-time Clusters of Crime Patterns in Houston (= Proc. of the Ninth Crime Mapping Research Conf., Pittsburgh PA), CD.
- LEITNER M., HELBICH M. (2009), Analyzing, Modeling, and Mapping the Impact of Hurricanes Katrina and Rita on the Spatial and Temporal Distribution of Crime in Houston, TX. In: CAR A., GRIESEBNER G., STROBL J. (Hrsg.), *Geospatial Crossroads @ GI\_Forum '09. Proc. of the Geoinformatics Forum Salzburg*, S. 110–119. Heidelberg, Wichmann.

- LEITNER M., HELBICH M. (2011), The Impact of Hurricanes on Crime: A Spatio-Temporal Analysis in the City of Houston, TX. In: *Cartography and Geographic Information Science*, 38, 2, S. 214–222.
- LEITNER M., BARNETT M., KENT J., BARNETT T. (2011), The Impact of Hurricane Katrina on Reported Crimes in Louisiana – A Spatial and Temporal Analysis. In: LEBEAU J.L., LEITNER M. (Hrsg.), *Spatial Methodologies for Studying Crime*. Special Issue of *The Professional Geographer*, 63, 2, S. 244–261.
- MILLER H., HAN J. (2009), *Geographical Data Mining and Knowledge Discovery*. Boca Raton, Taylor & Francis.
- NAKAYA T., YANO K. (2010), Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. In: *Transactions in GIS*, 14, 3, S. 223–239.
- OPENSHAW S., CHARLTON M., WYMER C., CRAFT A. (1987), A mark 1 geographical analysis machine for the automated analysis of point data sets. In: *Intern. Journal of Geogr. Information Science*, 1, S. 335–358.
- ROTH R.E., ROSS K.S., FINCH B.G., LUO W., MACEachREN A.M. (2010), A User-Centered Approach for Designing and Developing Spatiotemporal Crime Analysis Tools (= Proc. of GIScience 2010, Zürich), CD.
- SPIELMAN S.E., THILL J.-C. (2008), Social Area Analysis, Data Mining, and GIS. In: *Computers, Environment and Urban Systems*, 32, S. 110–122.
- SKUPIN A., HAGELMAN R. (2005), Visualizing Demographic Trajectories with Self-Organizing Maps. In: *GeoInformatica*, 9, S. 159–179.
- ULTSCH A. (1993), Self-Organizing Neural Networks for Visualization and Classification. In: OPITZ O., LAUSEN B., KLAR R. (Hrsg.), *Information and Classification*, S. 307–313. Berlin, Springer.
- VESANTO J. (1999), SOM-Based Data Visualization Methods. In: *Intelligent Data Analysis*, 3, S. 111–126.
- VESANTO J., ALHONIEMI E. (2000), Clustering of the Self-Organizing Map. In: *IEEE Transactions on Neural Networks*, 11, S. 586–600.