Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Konstruktion von SGB II-Dichten: Eine Möglichkeit sensible georeferenzierte pseudonymisierte Daten für die Wissenschaft aufzubereiten und für Analysen zur Verfügung zu stellen

Jörg-Peter Schräpler

(IT.NRW und Ruhr-Universität Bochum)

5. Konferenz des RatSWD Berlin, 14. Januar 2011

Gliederung

- FDZ: Zugang zu Mikrodaten aus dem Bereich der Amtlichen Statistik
- Georeferenzierte Daten und die Anonymisierungsproblematik
- Auswertung von Punktmustern
  - Beispiel BMBF-Projekt "Sozialraumindex für Schulen" und SGB II-Statistik der Bundesagentur für Arbeit
- Hintergrund des KDE-Verfahrens
- Anwendung im BMBF-Projekt
- Vorschlag für Ablauf und notwendige Maßnahmen im FDZ
- Fazit

Aufgaben der Forschungsdatenzentren

- Verbesserung der Dateninfrastruktur für wissenschaftliche Analysen
- Erleichterung des Zugangs der Wissenschaft zu amtlichen Mikrodaten

✓ Dienstleistungsangebot der amtlichen Statistik für die Wissenschaft wurde deutlich verbessert

Informationsgehalt und Anonymisierungsgrad

Einzeldaten	Informationsgehalt	Zugangsform
Nicht anonymisiert	Voller Informationsgehalt	Keine Zugangsmöglichkeit
Pseudonymisiert aber Angabe von Adressen	Keine Namen, aber Adressen liegen vor	Bisher keine Zugangsmöglichkeit
Formal anonymisiert	Keine direkten Identifikatoren (Name, Adresse) - Volles Auswertungspotential	Kontrollierte Datenfernverarbeitung
Faktisch anonymisiert	Information reduziert bis Aufwand der Deanonymisierung größer als potentieller Nutzen - Mittleres Auswertungspotential	Gastwissenschaftler Scientific Use Files
Absolut anonymisiert	Information reduziert bis Auskunftgeber nicht identifizierbar - niedriges Auswertungspotential	Public Use Files CAMPUS Files

Georeferenzierte Daten und Anonymisierungsproblematik

Weisen Einzeldaten raumbezogene Referenzinformationen auf, handelt es sich in der Regel um Daten, die aufgrund der Anonymisierungsproblematik interessierten Wissenschaftlern nicht zur Verfügung gestellt werden können,
z.B.

- pseudonymisierte Daten mit Adressangaben oder
- georeferenzierte Punktdaten.

→ Diese Raumdaten könnten allerdings für die Wissenschaft sehr wertvoll sein.

→ Von Interesse sind keine Einzelfälle, sondern i.d.R. Punktmuster bzw. die „Dichte“ der georeferenzierten Punktdaten.

Frage: Wie lassen sich aus sensiblen georeferenzierten Punktdaten über eine kDFV Dichteflächen zu erzeugen, die im Hinblick auf die Anonymisierung unbedenklich sind.

Auswertung von Punktmustern

z.B. die **räumliche Dichte** von:

- Hotspot-Analysen: registrierten Raubüberfällen
  (amerik. Justiz: „Crimestat“-Software)
- Krankheitsfällen (Epidemien)
- Tiervorkommen, Konstruktion von „Home Zones“ oder „Home Ranges“

Kriterien:

- Geokoordinaten müssen vorliegen.
- Das Untersuchungsgebiet muss vor der Analyse feststehen.

Georeferenzierte Daten

Beispiel:
Die Bundesagentur für Arbeit stellt den Statistischen Ämtern pseudonymisierte Adressdaten von SGB II-Empfänger(inne)n aus der Statistik der Grundsicherung für Arbeitsuchende nach § 53 SGB II zur Verfügung

Personendatensatz
Jeder Datensatz einer Person besteht aus folgenden Satzblöcken

Stand: November 2007

Satzblock-nummer	Satzblockbezeichnung	Beschreibung
P 1	Identifikatoren	Technische Informationen, die eine Verknüpfung der Datensätze erlauben.
P 2	Personenmerkmale	Soziodemografische Informationen über die Personen
P 3	Erwerbssituation	Erwerbsbezogene Informationen über die Personen
P 4	Bedarfe	Informationen zu den Bedarfen einer Person, aufgeschlüsselt nach den 25 Bedarfsarten
P 5	Einkommen	Einkommensbeträge nach Art der Einkommen (10 Arten), getrennt nach zu berücksichtigendem, verfügbarem und anrechenbarem Einkommen.
P 6	Sanktionen	Sanktionen nach §§ 30 und 31 SGB II nach Art und Umfang.
P 7	Leistung	Informationen zu den Leistungen (Ansprüchen) einer Person, aufgeschlüsselt nach den 25 Bedarfsarten
P 8	Dauer Leistungsbezug	Informationen über die Dauer des Leistungsbezuges in SGB II

SGBII - Daten

Personendaten P 2: Soziodemografie

Feld-Nr.	Merkmal	Ausprägung	Anmerkung / Beschreibung
P 2.1	Familienstand	ledig verheiratet verwitwet geschieden getrennt lebend eingetragene Lebenspartnerschaft eheähnliche Gemeinschaft keine Angabe Fehler im Ursprungswert keine Zuordnung möglich	Aufgrund unterschiedlicher Ausprägungen in den Quelldateien wurde ein Mappingverfahren durchgeführt, bei dem Daten aus Xsozial auf Ausprägungen aus A2LL codiert wurden.
P 2.2	Geschlecht	männlich weiblich keine Angabe Fehler im Ursprungswert keine Zuordnung möglich	
P 2.3	Alter	in Jahren	
P 2.4	Staatsangehörigkeit	nnn	dreistelliger Nationalitätenschlüssel; doppelte Staatsangehörigkeiten werden nicht erfasst, sollte deutsch eine der mehrfachen Staatsangehörigkeiten sein, so wird deutsch eingetragen.
P 2.6	Rolle in der Bedarfsgemeinschaft	BVEHB PTR MUK E-MUK E-MUK/PTR keine Angabe Fehler im Ursprungswert keine Zuordnung möglich	Verfahrensbedingte Kennzeichnung der BG-Mitglieder nach Stellung in der BG. Rolle ist nicht durch Gesetz definiert, sondern Ergebnis der Zusammenarbeit zwischen BG und Leistungsträger.

Georeferenzierte SGBII - Daten

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Adressangaben liegen vor:

Zur weiteren Auswertung werden die Adressen der Bedarfsempfänger georeferenziert

BMBF-Projekt:
Nutzung dieser Information zur Erfassung der „sozialen Belastung“ der Einzugsgebiete von Schulen in NRW

(Bild fiktives Beispiel)

BMBF-Projekt: Sozialindex für Schulen

Ziel: Erfassung der „Sozialen Belastung“ von Schulen durch Merkmale des Sozialraums auf Basis von regelmäßig anfallenden amtlichen Daten

- Zur konkreten Bestimmung der „sozialen Belastung“ muss das direkte Einzugsgebiet der Schule analysiert werden.
- So sollten in möglichst adäquater Weise die Einzugsgebiete und das Umfeld der Schulen in Hinblick auf vielfältige Dimensionen sozialer Ungleichheit beschrieben werden.

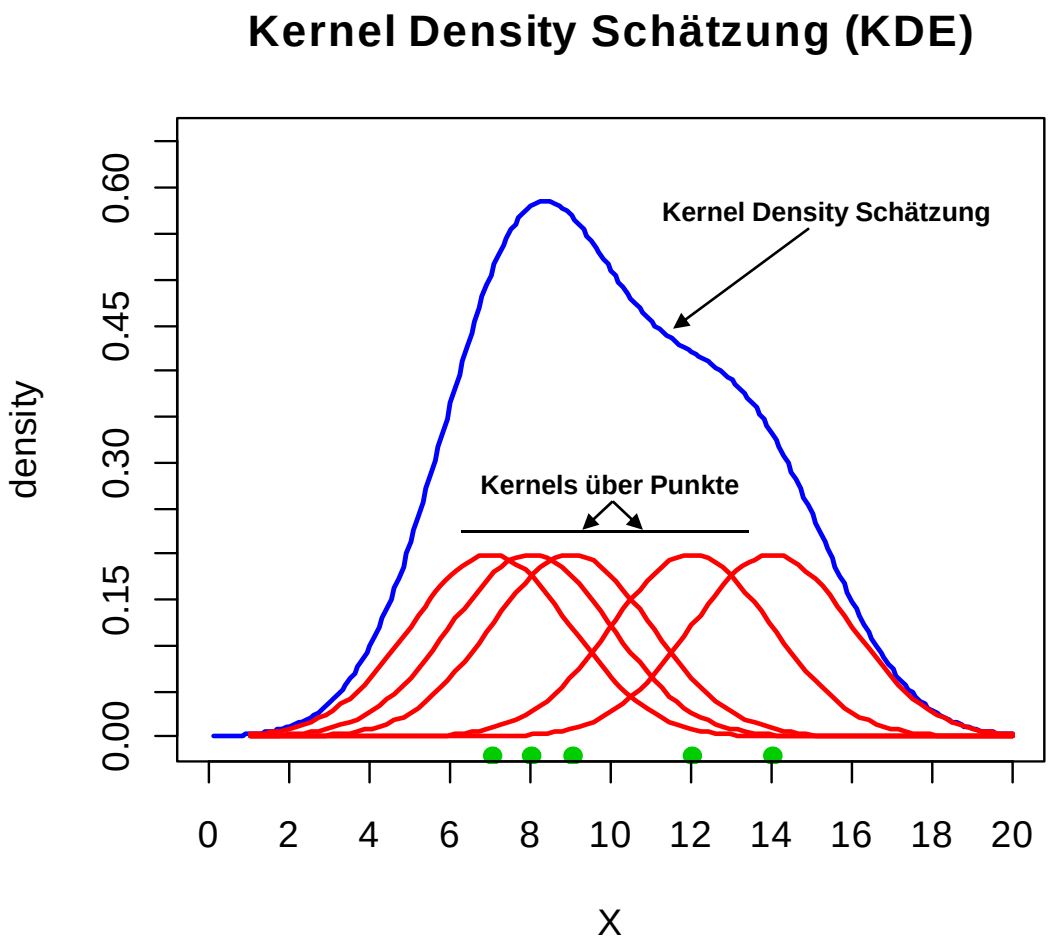
beteiligte Institute:
IT.NRW; ZEFIR; Bergische Univ. Wuppertal; IFS, Dortmund

Bonsen, Martin; Bos, Wilfried; Gröhlich, Carola; Harney, Benjamin; Imhäuser, Kurt; Makles, Anna; Schräpler, Jörg-Peter; Terpoorten, Tobias; Weishaupt, Horst; Wendt, Heike: Zur Konstruktion von Sozialindizes. *Bonn: BMBF (2010), 178 S.* .

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Hintergrund des KDE-Ansatzes

Was versteht man unter „Kernel-Density-Estimation“ ?

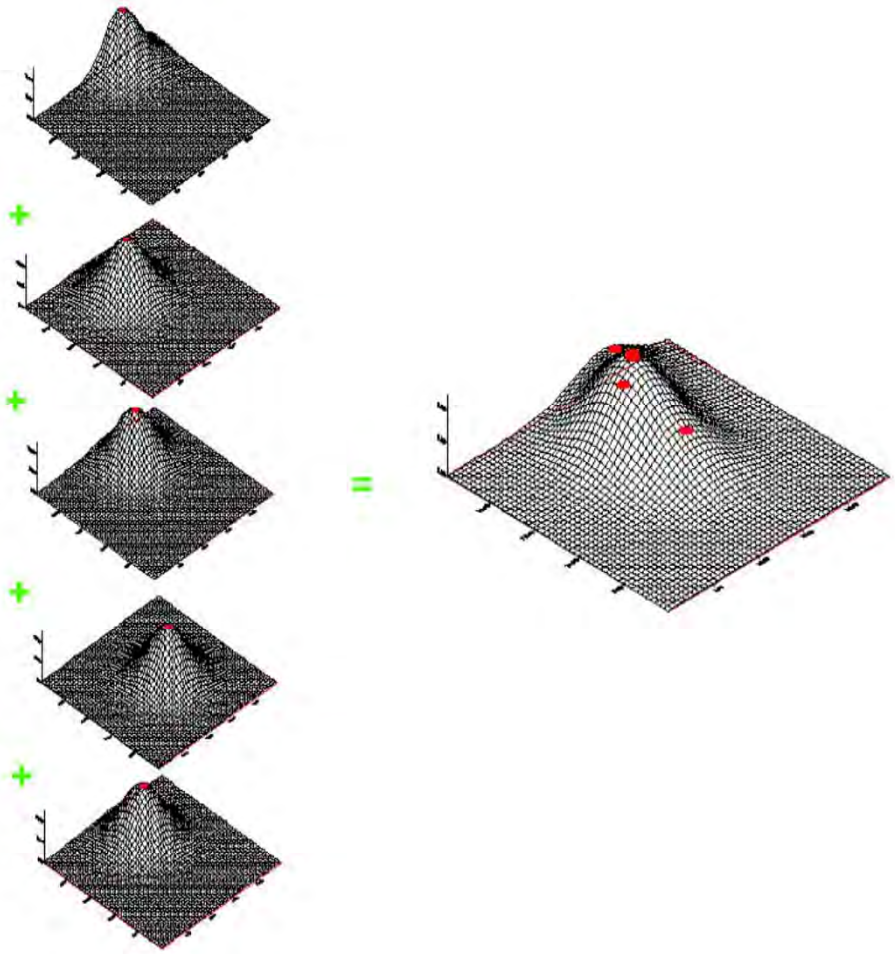


Aufsummierung von Normal-Kernels für 5 Punkte

Allgemein:
Eine statistische Methode zur Beschreibung der Dichte diskreter „Punkte“

- “Kernels” sind jene festzu-legenden Dichtefunktionen, welche über jeden einzelnen Punkt gesetzt werden
- Der Abstand zu einem Referenzpunkt wird auf Basis einer math. Funktion bewertet
- Die Werte der einzelnen „Kernels“ werden dann zu einer „Kernel Density“ aufsummiert.

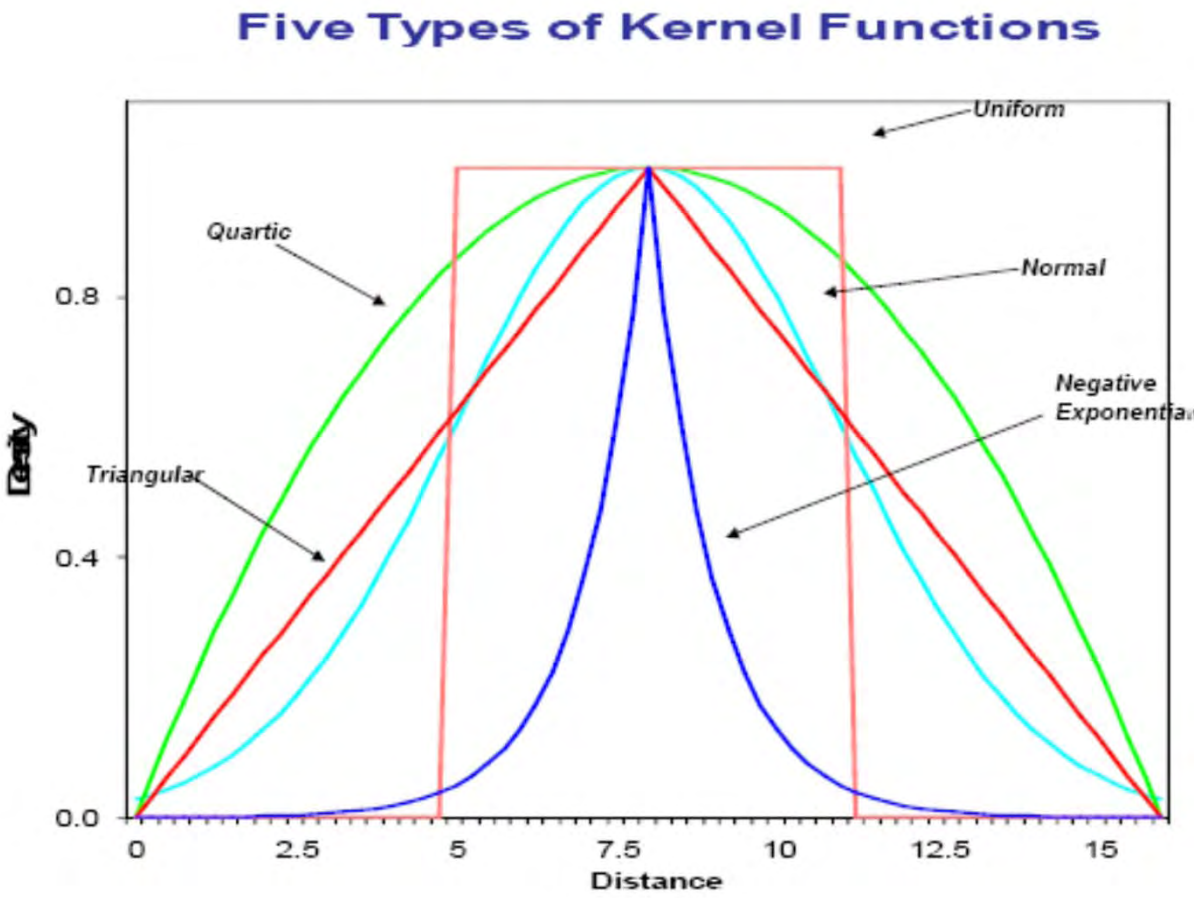
Bei vorliegender Geokoordinaten



1. **Die x- und y-Koordinaten bilden eine Fläche, auf die ein Gitter gelegt wird.**

2. **Die „Intensität“ oder „Risikowahrscheinlichkeit“ ergibt sich aus einem 3-dimensionalen Kernel**

Viele verschiedene Kernel-Funktionen sind möglich:



Beispiel: „normalverteilter Kernel“

- wird häufig verwendet
- Vorteil: Dichtefunktion nutzt Bereich ± ∞ und damit alle verfügbaren Punkte

Der Dichtewert an Gitterreferenzpunkt x_j ist dann (z.B. Levine 2004):

$$f(x_j) = \sum_{i=1}^{N} \left[W_i \cdot \frac{1}{h^2 2\pi} \cdot e^{-\left(\frac{d_{ij}^2}{2h^2}\right)} \right]$$

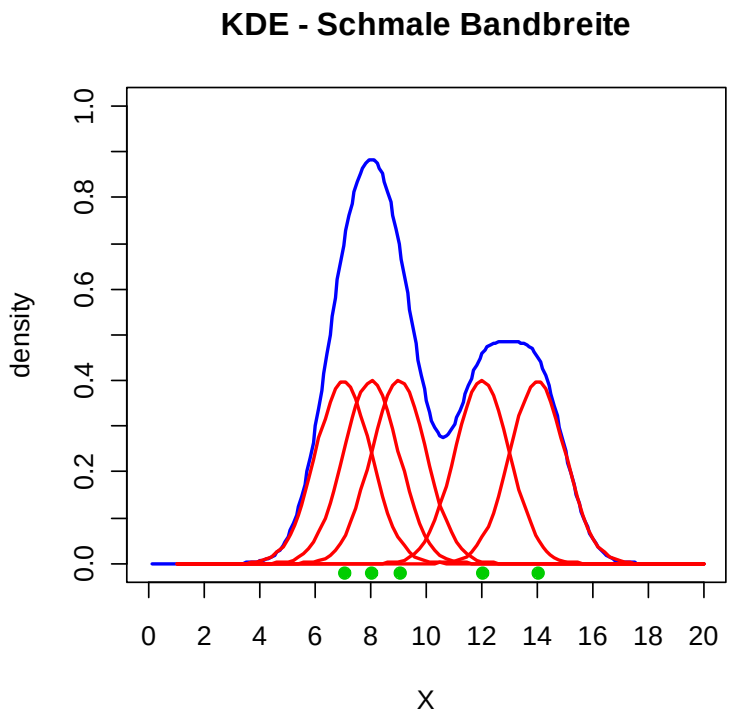
d_{ij} – Distanz zwischen dem jeweiligen Punkt x_i und dem Referenzpunkt x_j im Gitter (Gittermittelpunkt)

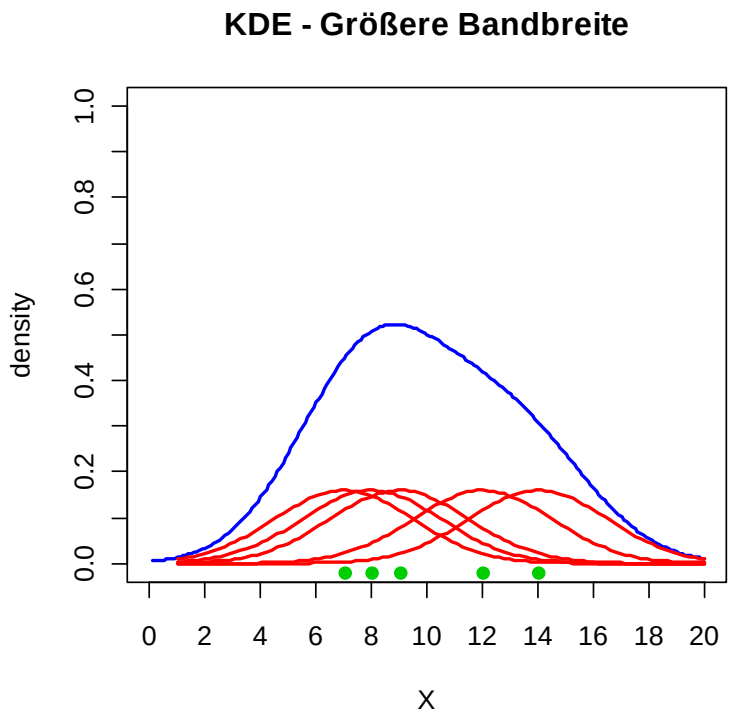
h - Bandwidth, bzw. Standardabw. der Normalverteilung

W_i – Gewichtungsvariable, falls vorhanden

Für alle Kernel gilt: Rand-Effekte (edge effects) können zu Verzerrungen führen

Die Wahl der Bandbreite beeinflusst das Ergebnis





- Die Glättung der Dichtefunktion ergibt sich dann aus der Bandbreite *h*

Wahl der Bandbreite

Um die Güte eines KDE Schätzers beurteilen zu können, ist es notwendig, ein Distanzmaß zwischen der wahren Dichte und der Schätzung zu definieren. Ein hierfür übliches Maß ist der so genannte *Integrated Squared Error* (ISE)

$$\boxed{ISE(x) = \int \left[\hat{f}_h(x) - f(x) \right]^2 dx}$$

Sein Erwartungswert ist der *Mean Integrated Squared Error* (MISE)

$$\boxed{MISE(x) = E\left[ISE(x)\right] = E\left[\int \left\{ \hat{f}_h(x) - f(x) \right\}^2 dx\right]}$$

wobei $\hat{f}_h(x)$ der KDE Schätzer bei gegebener Bandbreite *h* und *f(x)* der wahre Dichtewert ist. *MISE* ist ein Maß für die Güte des KDE Schätzers bei gegebener Bandbreite *h*.

Reference - Methode

Im Falle von normalverteilten Variablen und multivariat normalverteilten Kernels ist die optimale Bandbreite, die *MISE* asymptotisch minimiert durch h_{ref} gegeben (vgl. Bowman/Azzalini 1997, S. 32; Wand/Jones 1995, S. 111):

$$\boxed{h_{ref} = \left(\frac{4}{d+2}\right)^{1/(d+4)} n^{-1/(d+4)} \hat{\sigma}}$$

d - Dimension

bei zwei Dimensionen (x- und y-Koordinaten) gilt dann (Worton 1995):

$$\boxed{h_{ref} = \hat{\sigma} n^{-\frac{1}{6}}, \quad wobei \quad \hat{\sigma} = \sqrt{\frac{\hat{\sigma}_x^2 + \hat{\sigma}_y^2}{2}}}$$

Alternative:
LSCV – „Least-squares cross-validation“ Methode (Bowman 1984)

LSCV geht ebenfalls von *MISE* aus und nutzt die expandierte Form:

$$\boxed{MISE(x) = E\int\left[\hat{f}_h(x) - f(x)\right]^2 dx = E\int \hat{f}_h(x)^2 dx - 2E\int f(x)\hat{f}_h(x)dx + \int f(x)^2 dy}$$

Der letzte Term auf der rechten Seite hängt nicht von der Bandbreite h ab, die Minimierung von *MISE* ist somit äquivalent mit der Minimierung von

$$\boxed{MISE(x) - \int f(x)^2 dx = E\left[\int \hat{f}_h(x)^2 dx - 2\hat{f}_h(x) f(x)dx\right]}$$

Die rechte Seite ist unbekannt, da sie von der Dichte *f* abhängt. Der folgende Ausdruck ist hierfür jedoch ein unverzerrter Schätzer (vgl. Wand/Jones 1995, S.63)

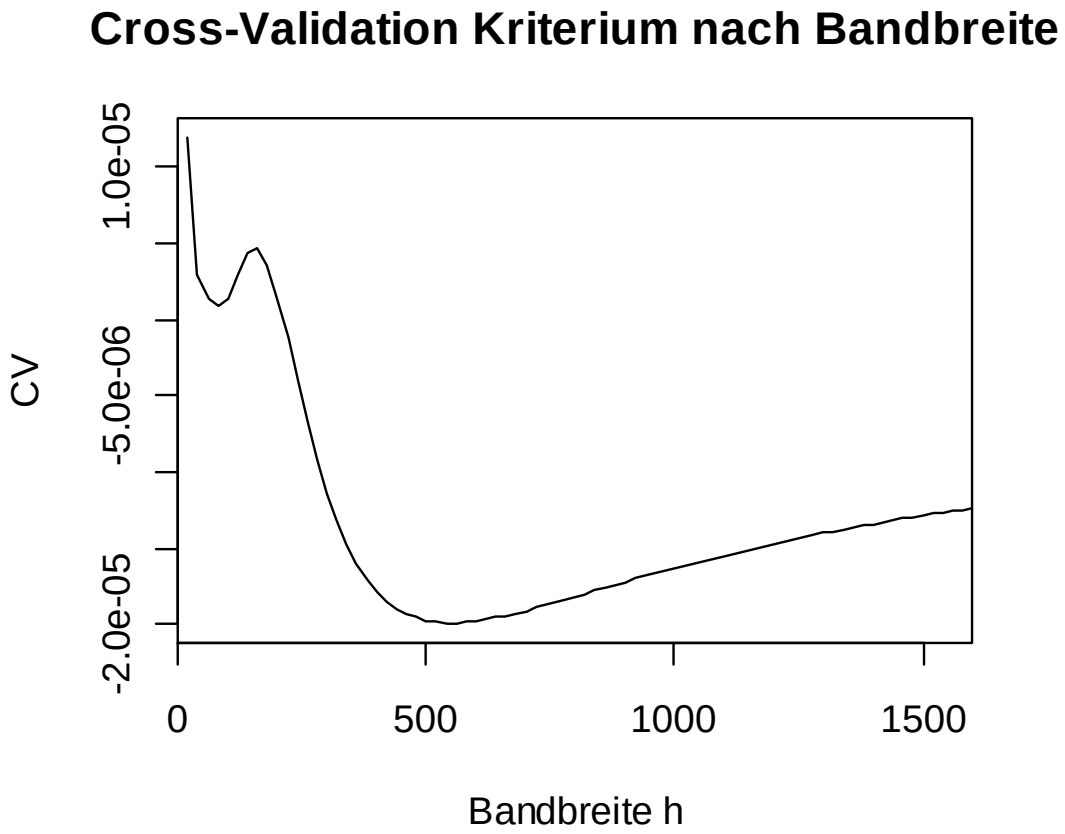
$$\boxed{LSCV(x) = \int \hat{f}_h(x)^2 dx - 2n^{-1}\sum_{i=1}^{n} \hat{f}_{h_{-1}}(x)}$$

wobei $\hat{f}_{h_{-1}}(x) = (n-1)^{-1}\sum_{j\neq i}^{n} K_h(x - X_j)$

bei normalverteilten (eindimensionalen) Kernel: $\hat{f}_{h_{-1}}(x) = (n-1)^{-1}\sum_{j\neq i}^{n} \phi(x - X_j; h)$

(Leave-one-out Kreuzvalidierung)

Das Least Square Cross Validation Kriterium in Abhängigkeit von der Bandbreite h – Beispiel Dortmund



Festlegung der Bandbreite – Beispiel Dortmund und Kreis Coesfeld

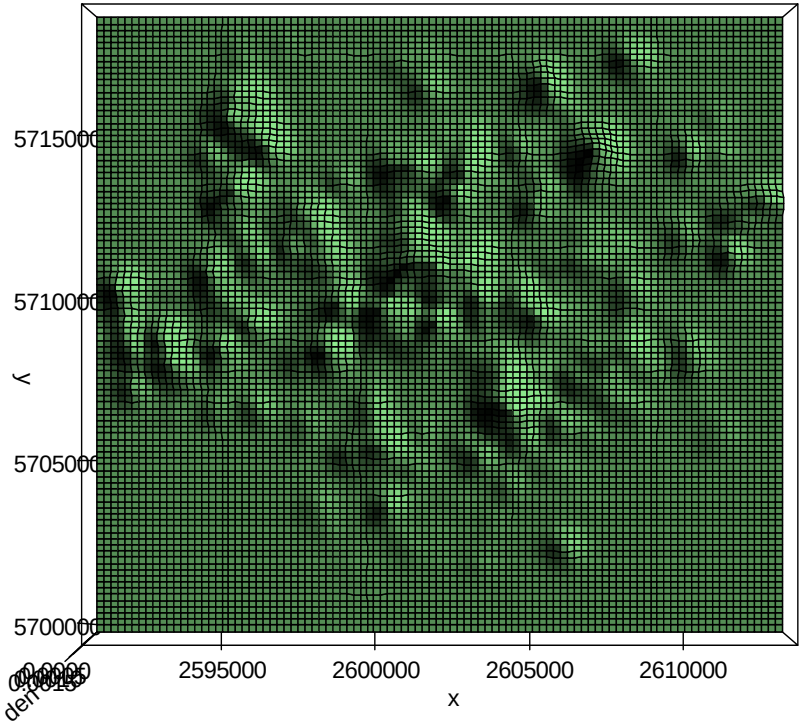
Tabelle 1: Berechnung der Bandbreiten für die KDE Schätzung in den Pilotregionen

Gemeinde	GKZ	N	%	$h_{ref}(x)$	h_{ref} (y)	h_{ref}	ε	p-Wert	CV(x)	CV(y)	CV
Ascheberg	5558004	259	0,84	854,0	1.286,7	1.092,0	20,2	0,0	471,7	710,6	605,9
Billerbeck, Stadt	5558008	135	0,44	288,2	226,4	259,2	8,3	0,0	279,7	219,8	255,4
Coesfeld, Stadt	5558012	520	1,68	349,5	503,8	433,6	48,6	0,0	413,8	596,5	518,4
Dülmen, Stadt	5558016	707	2,28	931,3	632,6	796,1	83,1	0,0	944,8	641,8	787,1
Havixbeck	5558020	160	0,52	327,8	361,5	345,1	13,9	0,0	364,1	401,6	381,4
Lüdinghausen, Stadt	5558024	339	1,09	628,6	265,5	482,5	19,1	0,0	612,0	258,4	159,1
Nordkirchen	5558028	81	0,26	1.044,2	675,3	879,3	9,7	0,0	135,8	87,8	111,8
Nottuln	5558032	303	0,98	1.108,2	603,0	892,1	19,6	0,0	494,9	269,3	374,7
Olfen, Stadt	5558036	129	0,42	490,2	362,2	431,0	20,8	0,0	427,0	315,5	367,8
Rosendahl	5558040	107	0,35	1.695,9	608,8	1.274,1	8,0	0,0	724,5	260,1	351,8
Senden	5558044	366	1,18	636,3	648,5	642,4	46,8	0,0	516,5	526,4	521,4
Dortmund, Stadt	5913000	27902	89,98	841,1	536,1	705,3	300,1	0,0	698,4	445,2	550,3
arithm. Mittel				**766,3**	**559,2**	**686,1**			**506,9**	**394,4**	**415,4**
Summe		**31.008**	**100**								

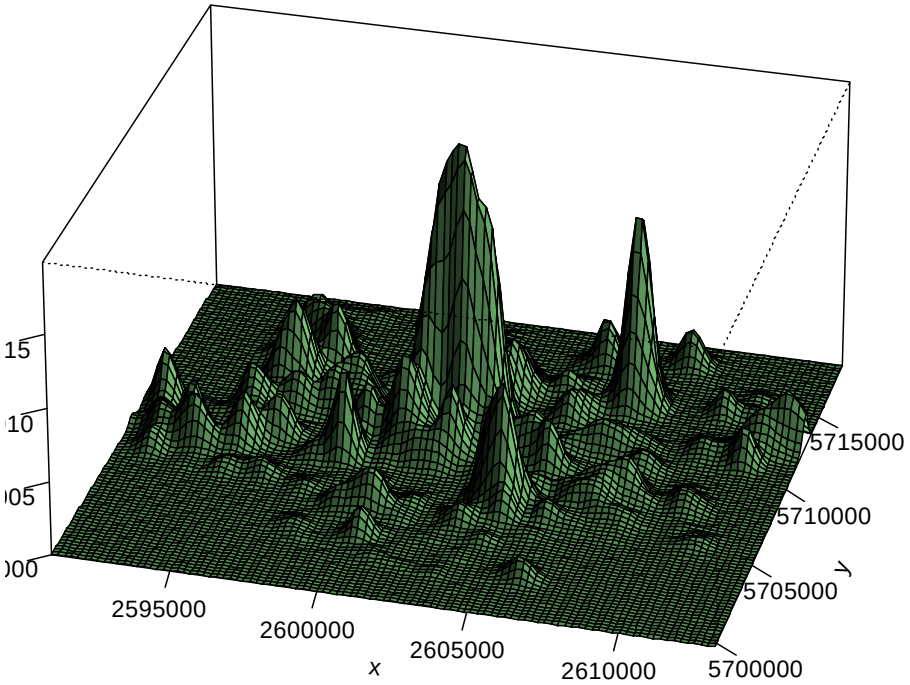
Quelle: eigene Berechnungen, SGBII-Empfängerdatei (SGBII-Empfänger ≤ 18 Jahre)

Kernel Density Surface für Dortmund

SGBII-Dichte in Dortmund

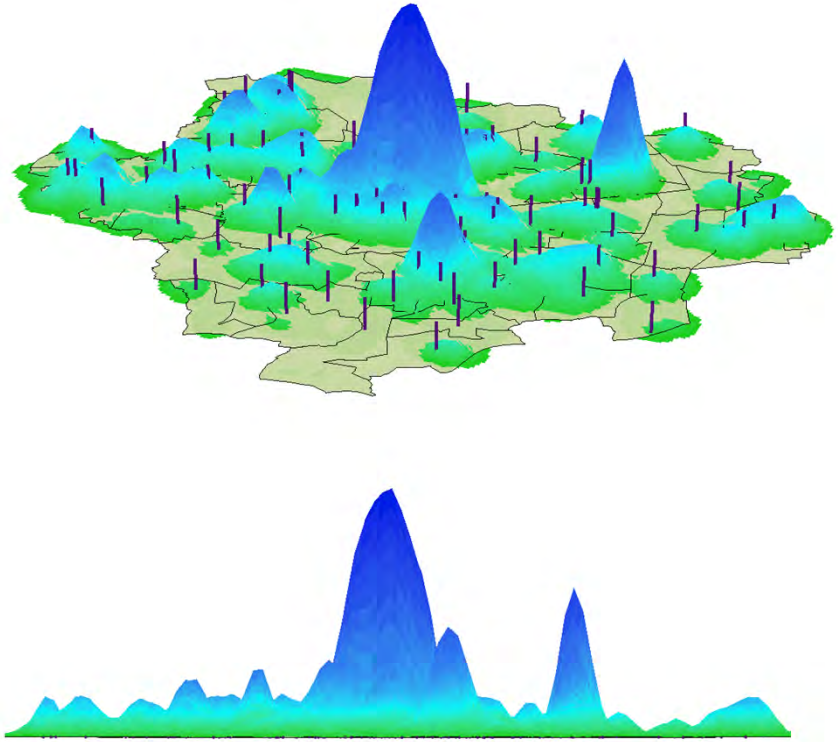


SGBII-Dichte in Dortmund



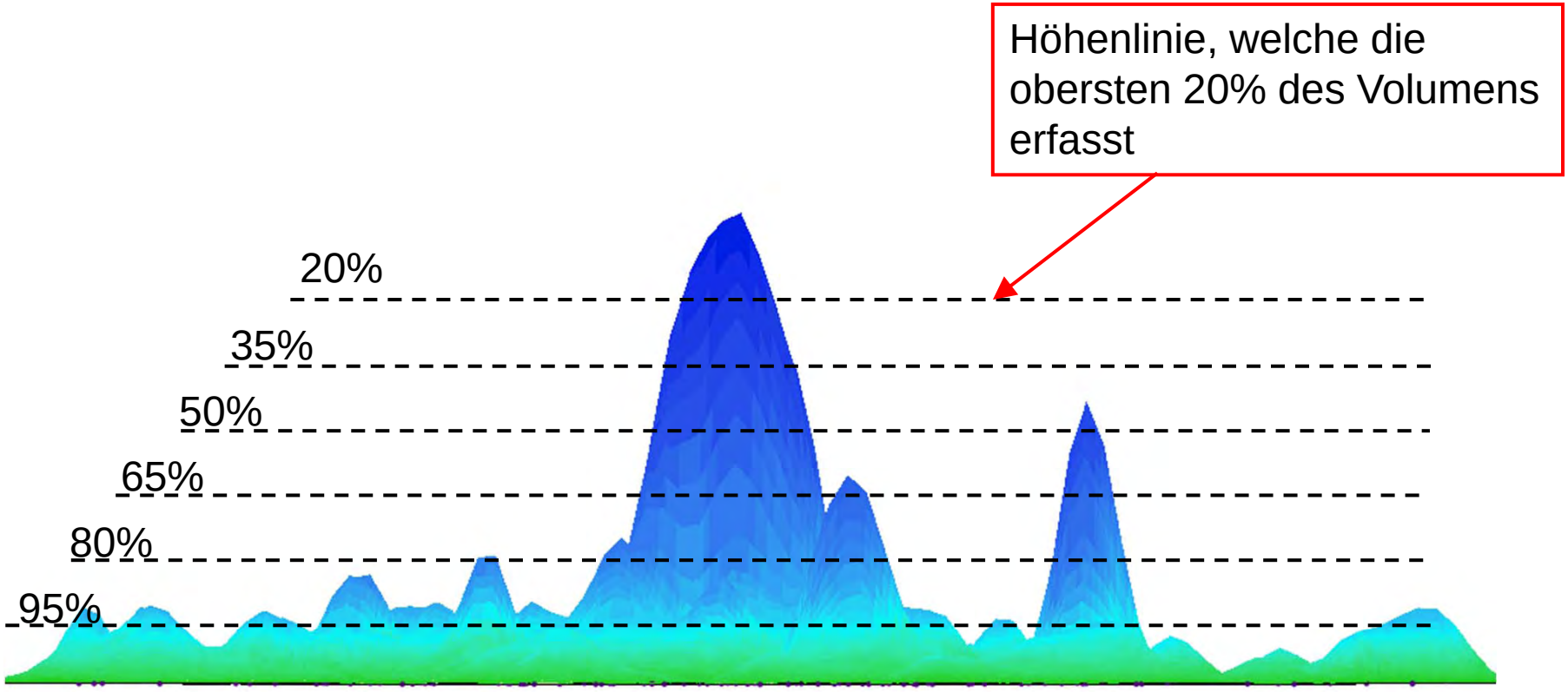
3D- Fläche der SGBII-Dichte mit Schulstandorten

Kreis Coesfeld



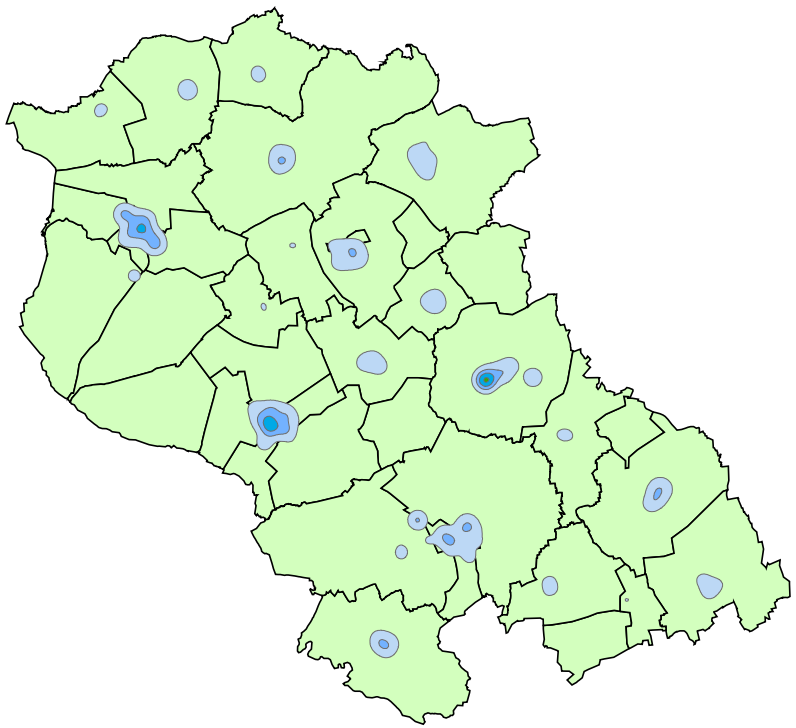
Dortmund

Percentage Volume Contours

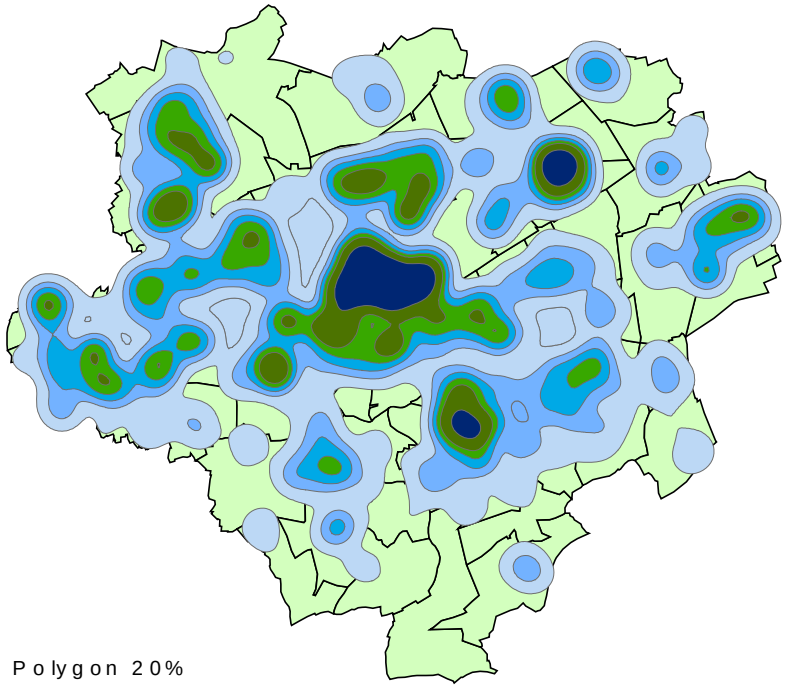


Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Kernel Density Schätzungen für SGBII – Empfänger (≤ 18 Jahre)
Percentage Volume Contour (PVC) Plots
(20% / 35% / 50% / 65% / 80% / 95%)
(Bandwidth 500)



Kreis Coesfeld



Dortmund

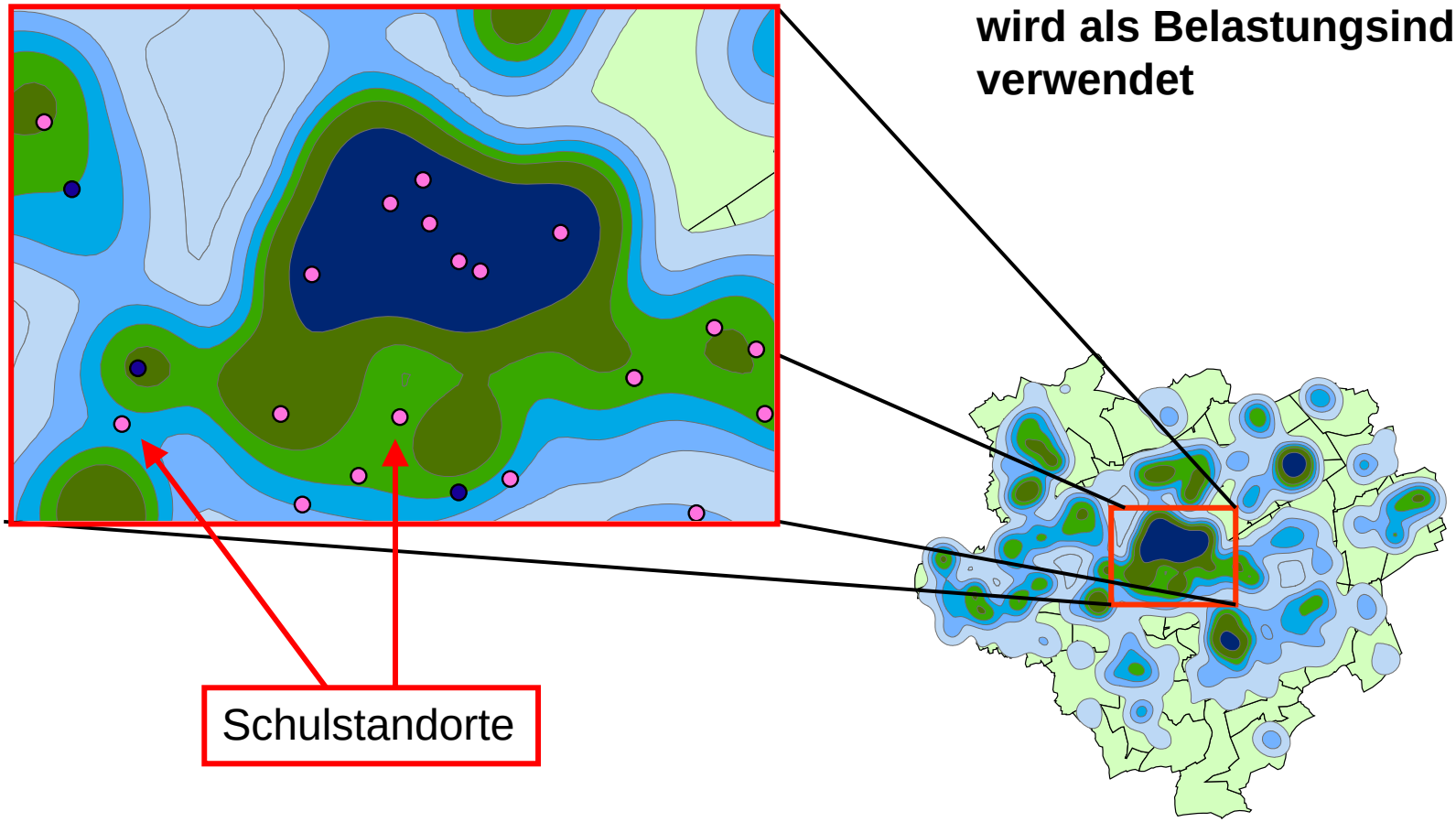
Legende

- KDE Polygon 20%
- KDE Polygon 35%
- KDE Polygon 50%
- KDE Polygon 65%
- KDE Polygon 80%
- KDE Polygon 95%

KDE- Ansatz Wiesbaden, 14.01.2011

PVC – Ausschnitt für Dortmund

Schulstandort in dem jeweiligen Dichtebereich wird als Belastungsindikator verwendet



Klassierte KDE-Variable

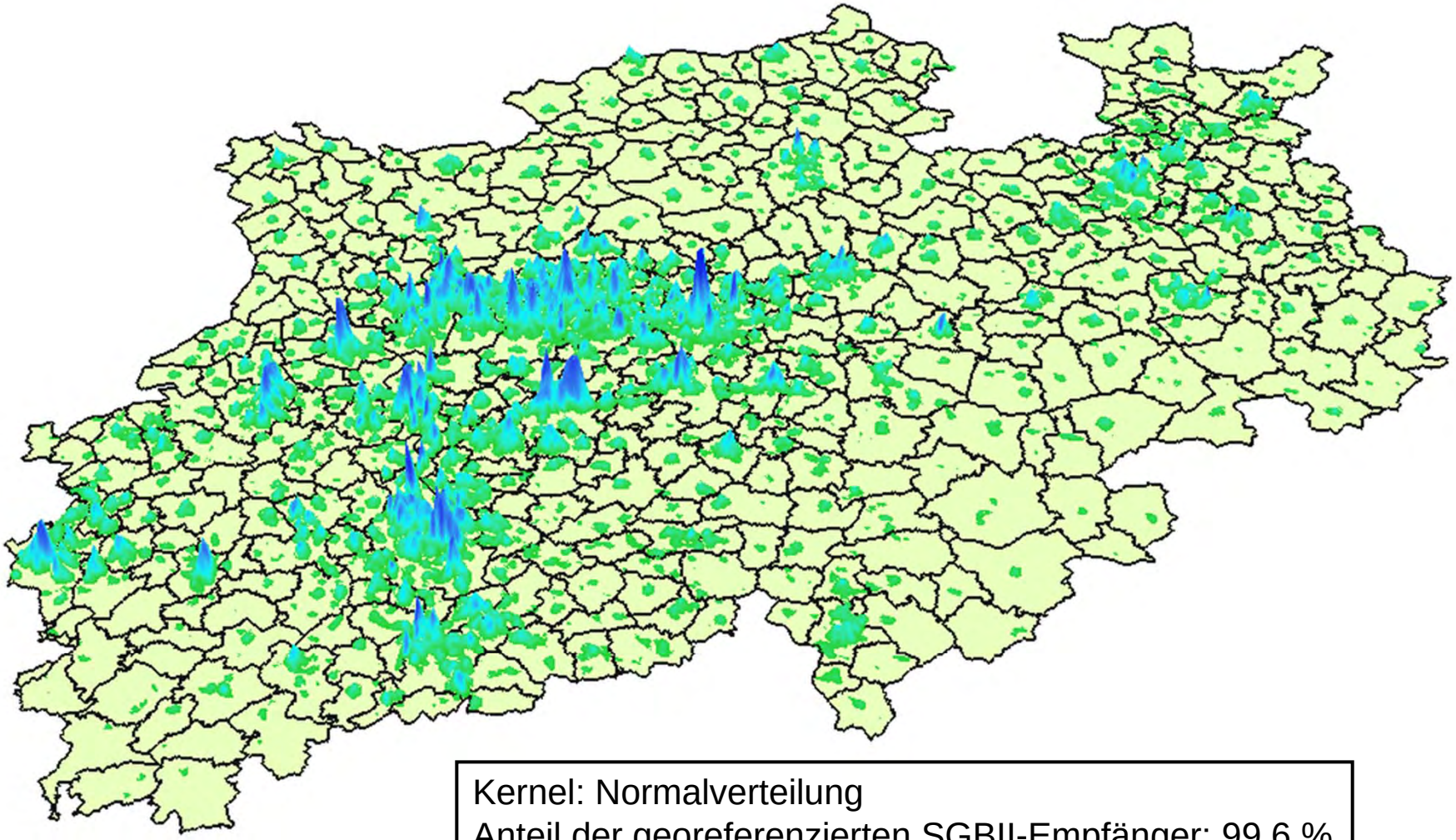
Klassifizierung von SGBII - Dichten

Percentage Volume Contour	***Kodierung***	***Dichte***
≤ 20 %	1	sehr hohe SGBII-Dichte
> 20 – 35%	2	hohe SGBII-Dichte
> 35 – 50%	3	mittelstarke bis hohe SGBII-Dichte
> 50 – 65%	4	mittelstarke SGBII-Dichte
> 65 – 80%	5	mittelstarke bis niedrige SGBII-Dichte
> 80 – 95%	6	niedrige SGBII-Dichte
> 95%	7	sehr niedrige SGBII – Dichte

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Anwendung auf ganz Nordrhein-Westfalen

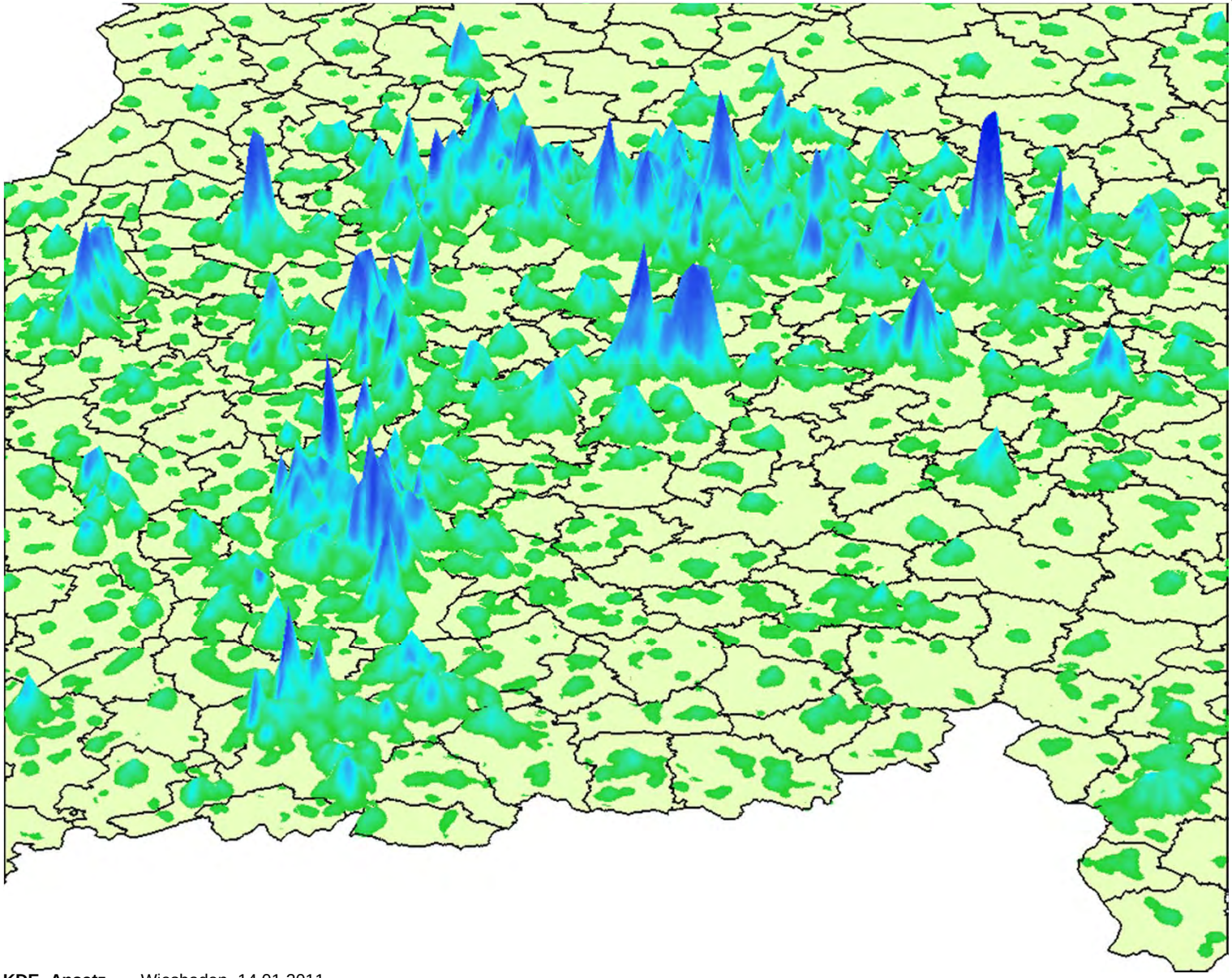
SGBII-Dichteverteilung für NRW

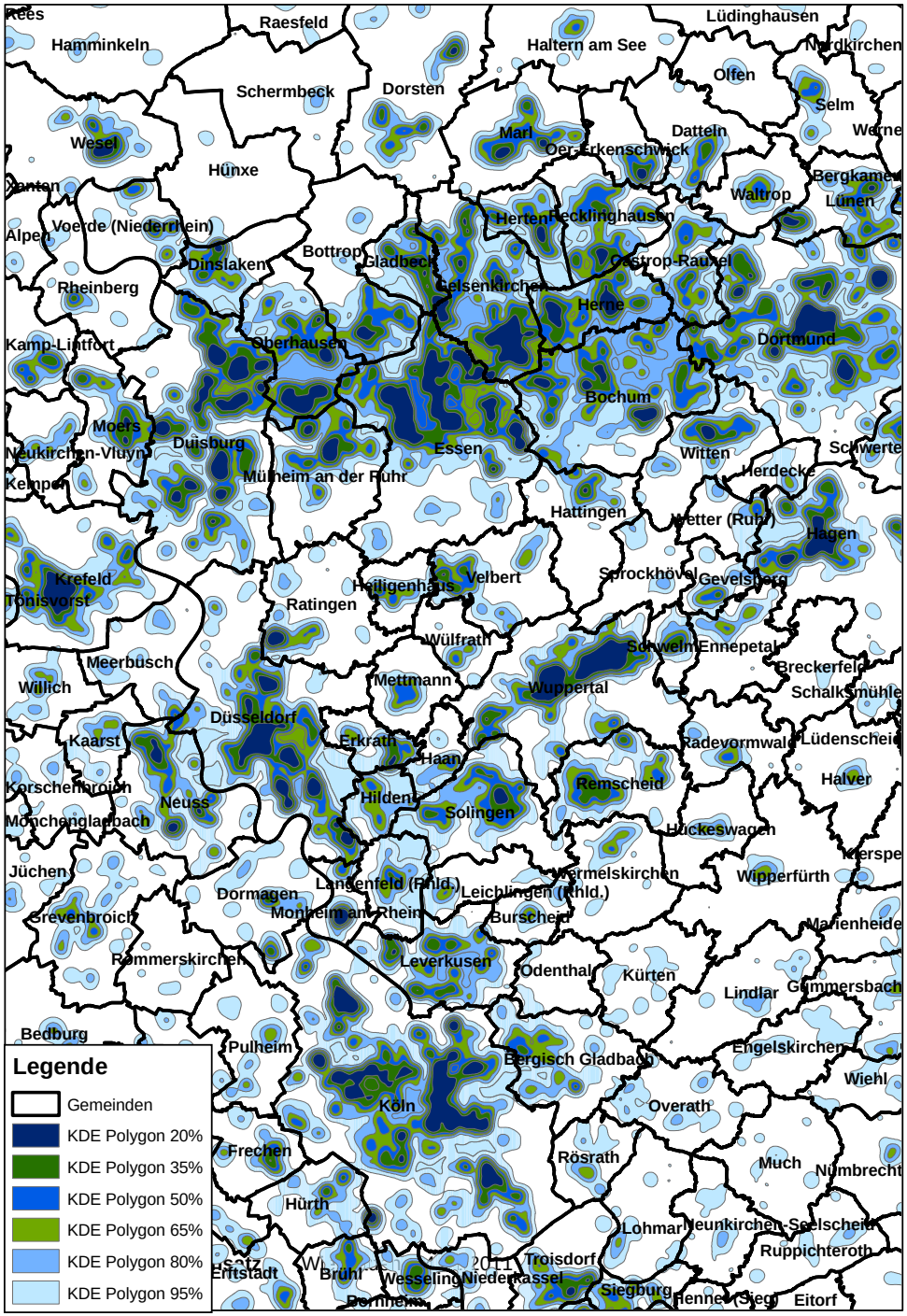


Kernel: Normalverteilung
Anteil der georeferenzierten SGBII-Empfänger: 99,6 %
Anzahl der SGBII-Empfänger ≤ 18 Jahre: 574.486

SGBII-Dichteverteilung (Ausschnitt)

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik





Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Percentage Volume Contours

-Klassifizierung von SGBII - Dichten und Darstellung als Flächen

Percentage Volume Contour	*Kodierung*	*Dichte*
≤ 20 %	1	sehr hohe SGBII-Dichte
> 20 – 35%	2	hohe SGBII-Dichte
> 35 – 50%	3	mittelstarke bis hohe SGBII-Dichte
> 50 – 65%	4	mittelstarke SGBII-Dichte
> 65 – 80%	5	mittelstarke bis niedrige SGBII-Dichte
> 80 – 95%	6	niedrige SGBII-Dichte
> 95%	7	sehr niedrige SGBII – Dichte

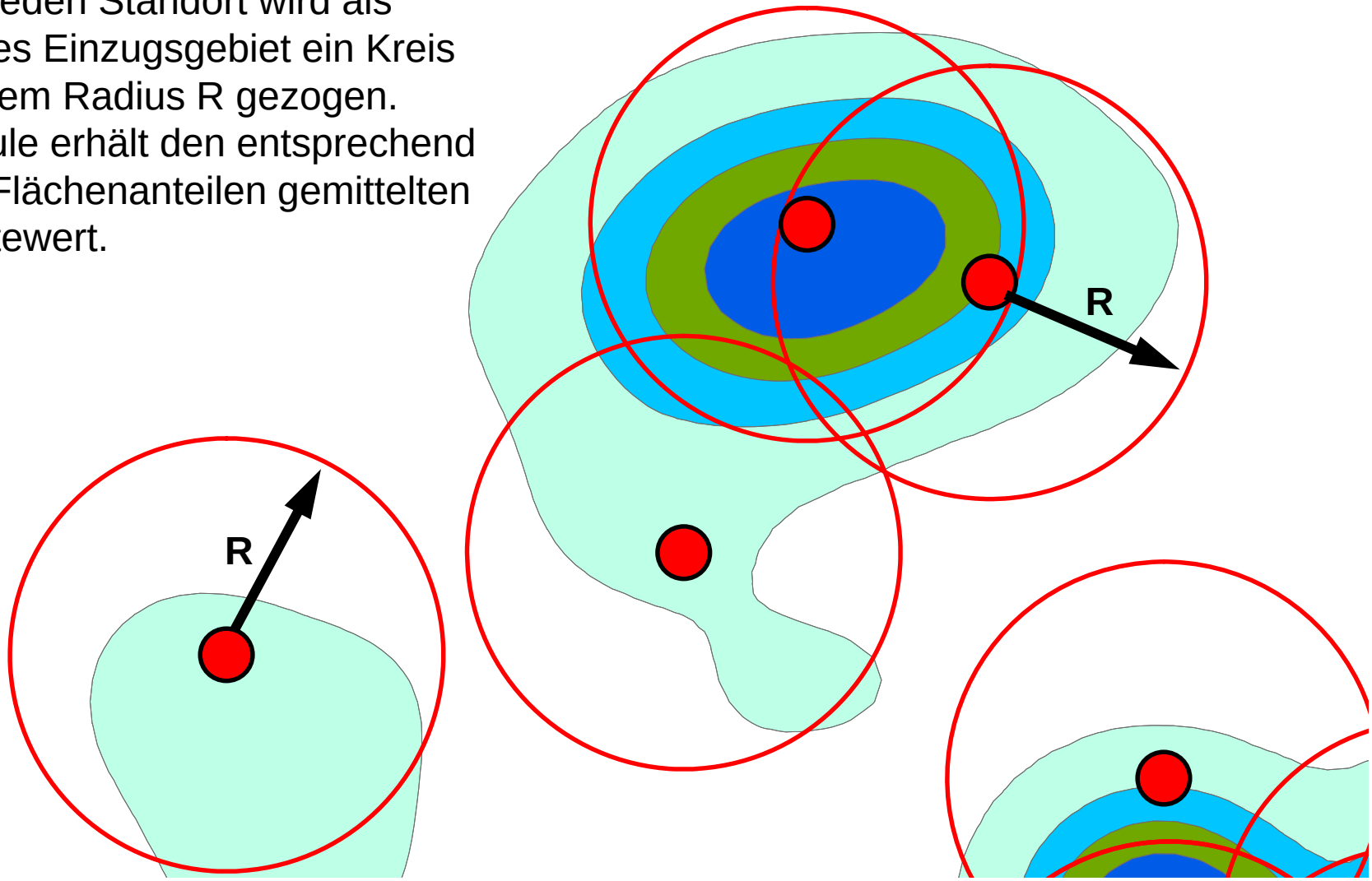
Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

Bildung eines Raumindex für Grundschulen und weiterführende Schulen in NRW

KDE- Ansatz Wiesbaden, 14.01.2011

Konstruktion von fiktive Einzugsgebieten

- Um jeden Standort wird als fiktives Einzugsgebiet ein Kreis mit dem Radius R gezogen.
- Schule erhält den entsprechend den Flächenanteilen gemittelten Dichtewert.



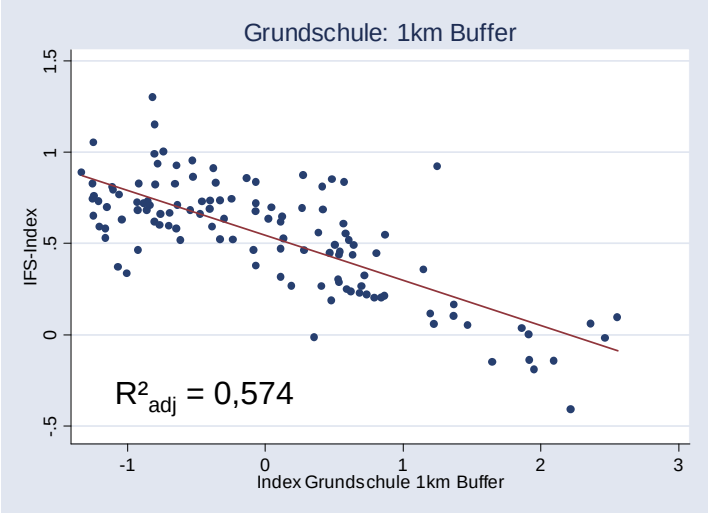
Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

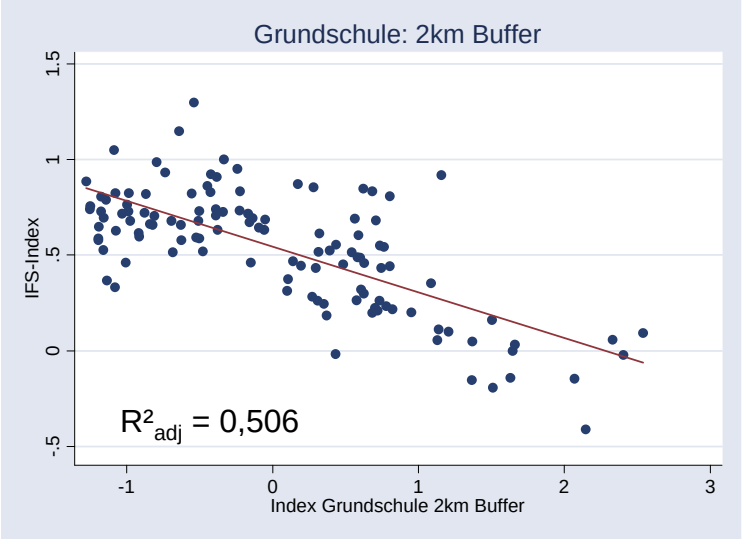
Evaluierung

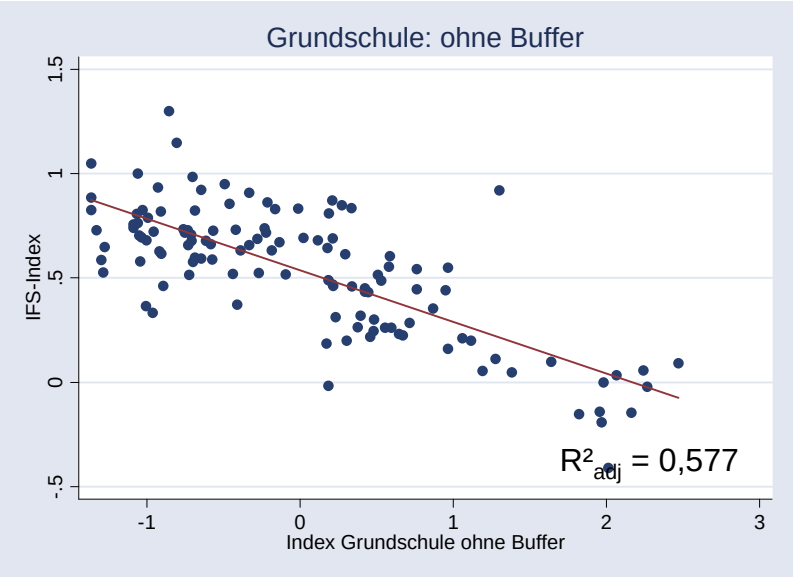
Index Grundschulen u. IFS-Index

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

nur Dortmund und Kreis Coesfeld







Erklärte Varianz des IFS-Index

	Buffer-Radius		
Grundschule	**ohne**	**1 km**	**2 km**
KDE Variable	0,477	0,474	0,327
Belastungsindex	0,577	0,574	0,506

Weiterführende Schulen – erklärte Varianz bei Lernstandserhebungen

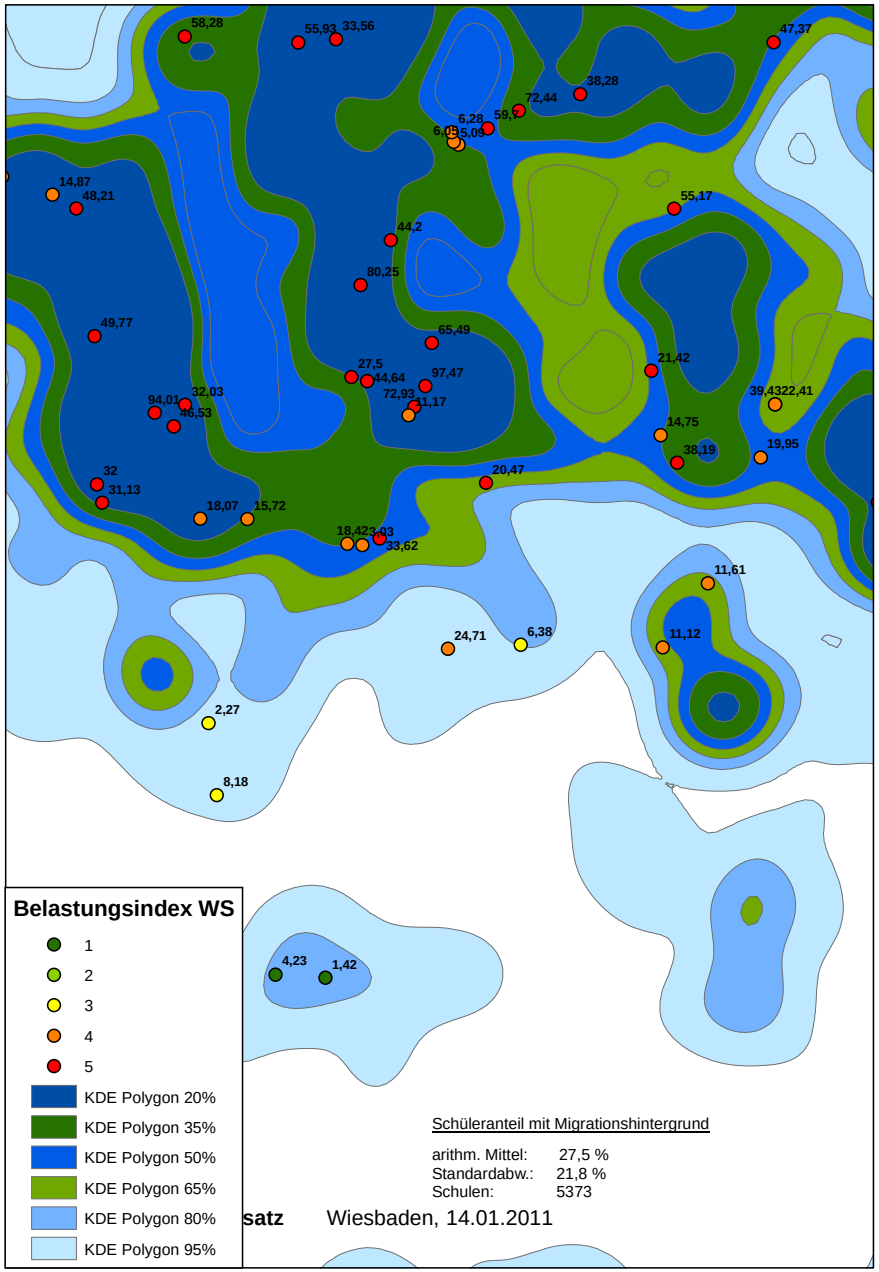
Erklärte Varianz der Lernstandserhebung Mathematik 8. Klasse durch die Belastungsindizes (**KDE-Variable + Migrationsstatus**) in Abhängigkeit von den Buffer-Radien

		Einzugsgebiete – Buffer-Radius				
Schulform	**ohne**	**1 km**	**2 km**	**3 km**	**4 km**	**5 km**
Hauptschule	0,362	0,370	0,390	0,411	0,431	0,446
Realschule	0,252	0,268	0,294	0,314	0,331	0,342
Gesamtschule	0,271	0,305	0,320	0,313	0,320	0,330
Gymnasium	0,165	0,166	0,159	0,163	0,169	0,174
Zusammen	0,218	0,193	0,185	0,195	0,204	0,213

Beispiel Stadt Essen

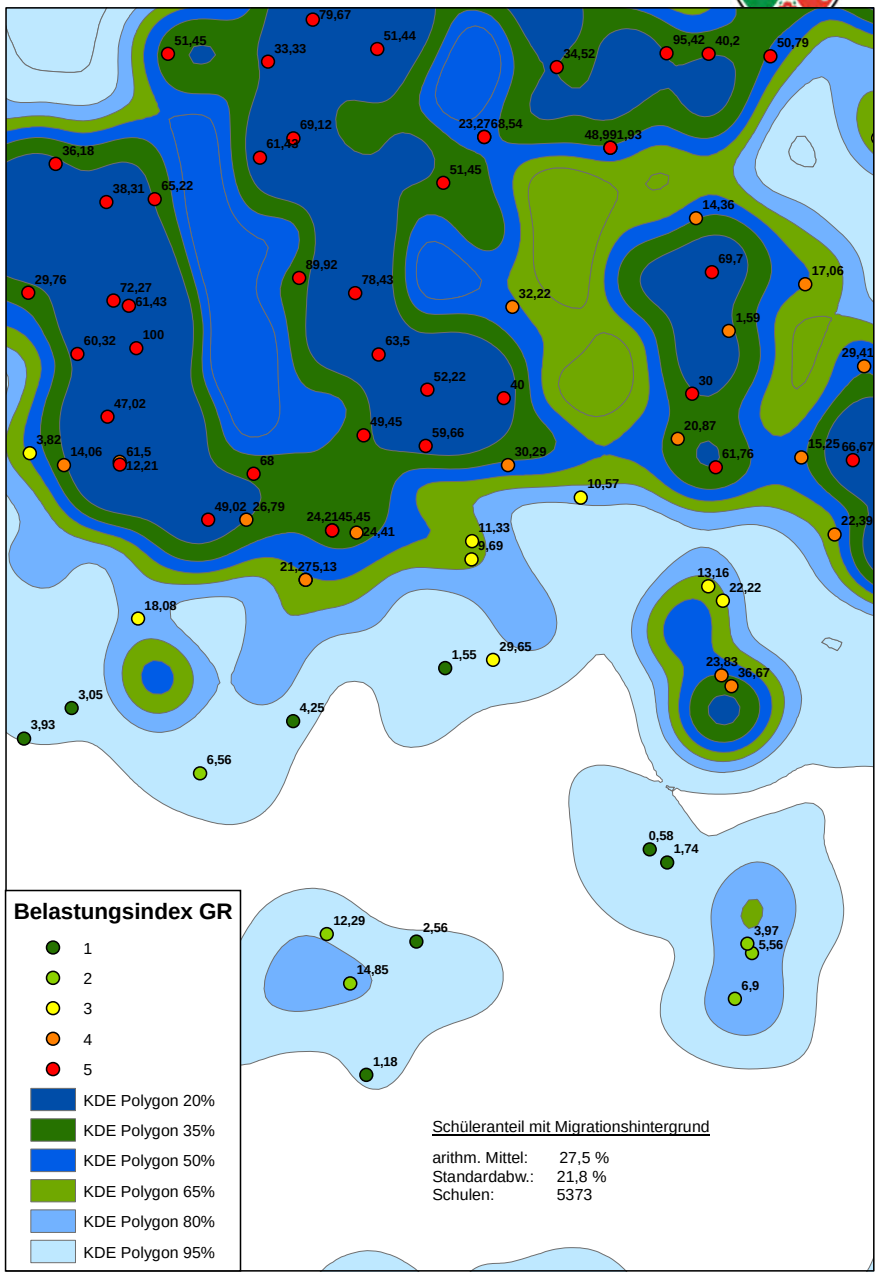
Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

weiterführende Schulen



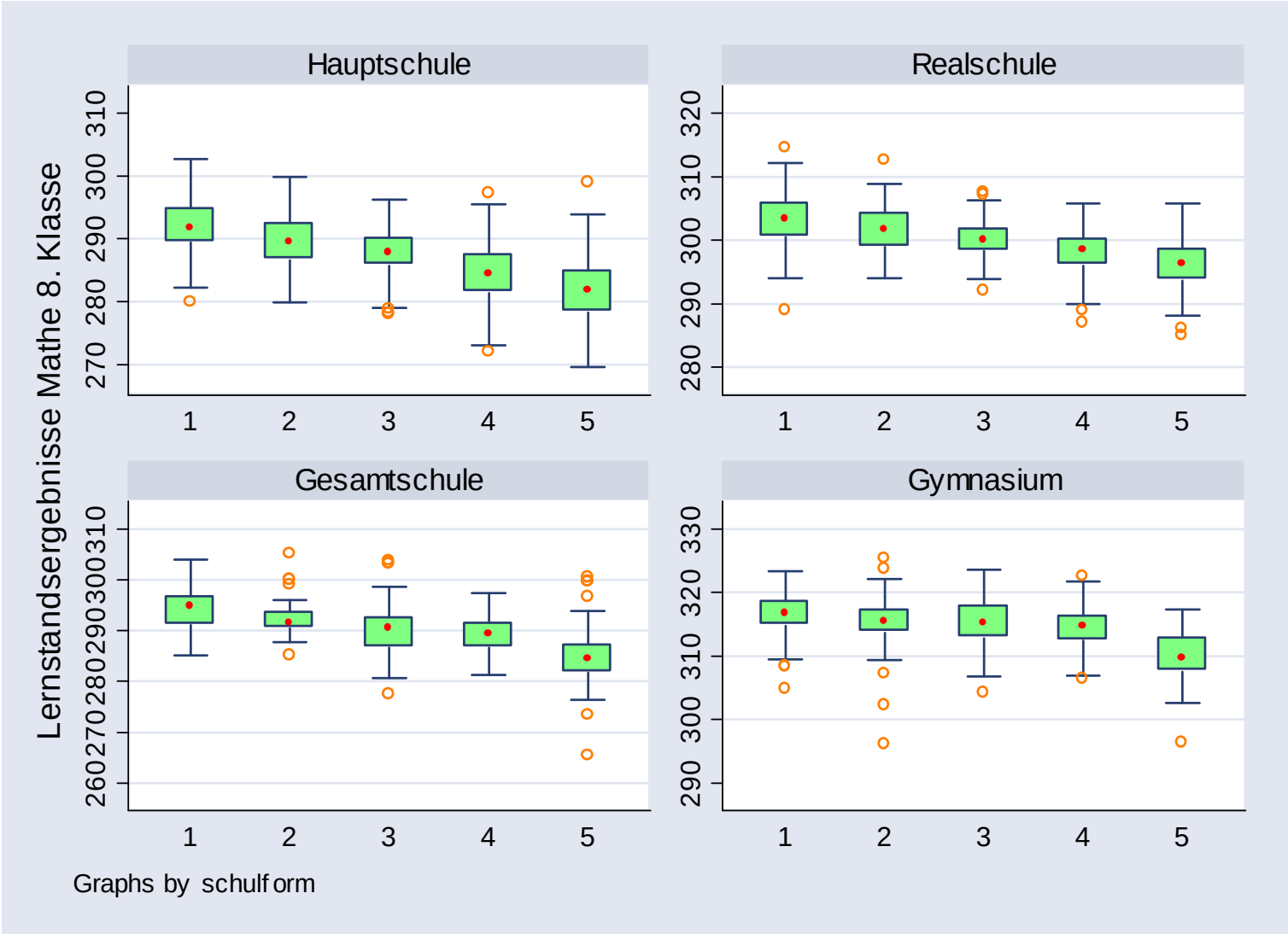
satz Wiesbaden, 14.01.2011

Grundschulen



Lernstandsergebnisse LSE8 nach Schulformen u. Belastungsstufen (Quintile)

Information und Technik Nordrhein-Westfalen
Geschäftsbereich Statistik



Notwendige Maßnahmen

1. **Rechtliche Prüfung:**
   a) Zugang zu der SGB II-Statistik der BA über eine kontrollierte Datenfernverarbeitung (kDFV) in den Statistischen Ämtern oder direkt bei der Bundesagentur für Arbeit.

2. **Maßnahmen in den FDZ der Statistischen Ämter:**
   a) Es müssen Rechner mit Geoinformationssystemen (und evtl. der Software R) ausgestattet und u.U. für Fernrechnen (Ausführen von Auswertungsprogrammen) vorbereitet werden.
   b) Betreuungspersonal muss mit GIS-Programmen vertraut sein.
   c) Sicherstellung, dass keine Bildschirmdarstellung mit georeferenzierten Punktdaten erfolgt.

3. **Wissenschaftler:**
   a) Schreiben von Software-Programmen zur Bestimmung der Bandbreite.
   b) Vorgabe der Percentage-Volume-Contour Abstufungen.

(möglicher) Ablauf für ganz NRW über kDFV

1. **Georeferenzierung in den Statistischen Ämtern bzw. FDZ:**
   die SGBII-Empfängeradressen (ca. 1,7 Millionen) und Schuladressen müssen georeferenziert werden.
   a) Für den Schulindex werden nur die Koordinaten der ≤ 18 jährigen Bedarfsempfänger verwendet (N = 574.486).

2. **Remote-Access:**
   Festlegung der Kernelfunktion und der Bandbreite
   a) Über kDFV muss ein Zugang zu den georeferenzierten Punktdaten gewährleistet sein um mittels eines geeigneten Programms (z.B. das „sm“ Package von Bowman/ Azzalini) die optimale Bandbreite zu bestimmen.

3. **GIS Verarbeitung über Remote-Access oder Vorgabe:**
   a) Erzeugung von „Percentage Volume-Contour“ (PVC) Linienplots.
   b) Die PVC-Plots werden in Polygon-Flächen umgewandelt.
   c) Optional für Schulindex: im letzten Schritt wird geprüft, in welchem Dichtebereich der Schulstandort liegt.

4. **Weitergabe des Outputs (Shape-Datei): Karte mit den Dichtepolygonen**

Fazit

- Mit Hilfe des KDE Verfahrens ist es möglich, aus sensiblen georeferenzierten Punktdaten, wie den pseudonymisierten Adressdaten von SGB II-Empfängern der Bundesagentur für Arbeit, Dichteflächen zu erzeugen, die im Hinblick auf die Anonymisierung unbedenklich sind.

- Die Auswertung von Punktdaten wie SGBII-Adressen bietet für die Sozialberichterstattung vielfältige neue Analysemöglichkeiten:

  - Unabhängig von Gemeinde- oder Stadtteilgrenzen können Problemgebiete viel genauer lokalisiert werden.

  - Die erzeugten Dichteflächen lassen sich mit anderen Raumdaten (stadtteilspezifische Daten) oder Punktdaten (z.B. Schulstandorte, Standorten von Krankenhäusern) sinnvoll verknüpfen.

- Die kontrollierte Datenfernverarbeitung im FDZ sorgt dafür, dass die georeferenzierten Adressdaten in den Statistischen Ämtern verbleiben und der Analysezugriff und die Kartenerstellung kontrolliert wird.

Software

Für die 3D-Grafiken und Konstruktion der Percentage Volume Contours:

- *ArcGIS*
  ESRI (2005) ArcGIS 9.2. Im Internet unter: http://www.esri.com
- Beyer, Hawthorne L. (2004) Hawth's Analysis Tools for ArcGIS. Im Internet unter: http://www.spatialecology.com/htools.

Zur Berechnung der Bandbreite:

- *CrimeStat*
  Levine, N. (2004) CrimeStat III Version 3.0. A Spatial Statistics Program for the Analysis of Crime Incident Locations. Houston, TX. The National Institute of Justice, Washington, DC.
- *R – SM Package*
  Bowman, A. W. and Azzalini, A. (2007). sm: kernel smoothing methods: Bowman and Azzalini (1997). R package version 2.2.

Literatur

Übersicht:

- Levine, N. (2004) CrimeStat III Version 3.0. A Spatial Statistics Program for the Analysis of Crime Incident Locations. Houston, TX. The National Institute of Justice, Washington, DC.
- Sheather, S. J. (2004) Density Estimation. Statistical Science, Band 19, Nr.4, S. 588-597.
- Silverman, B. W. (1986) Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
- Turlach, B.A., (1993) Bandwidth selection in kernel density estimation: A review, *Discussion Paper 9307*, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- Wand, M. P. und M. Chis Jones (1995) Kernel Smoothing. Chapman & Hall, London.

Anwendung:

- Schräpler, J.-P. (2009) Verwendung von SGBII - Dichten als Raumindikator für die Sozialberichterstattung am Beispiel der „sozialen Belastung“ von Schulstandorten in NRW - ein Kernel Density Ansatz. Statistische Analysen und Studien NRW, Band 57, p. 3-28.
- Vokoun, J. C. (2003) Kernel Density Estimates of Linear Home Ranges for Stream Fishes: Advantages and data Requirements. North Amrican Journal of Fisheries Management, 23, S. 1020 – 1029.
- Worton B.J. (1995) Using Monte Carlo simulation to evaluate kernel-based home range estimates. Journal of Wildlife Management 59, S.794-800.

Vielen Dank für Ihre Aufmerksamkeit !

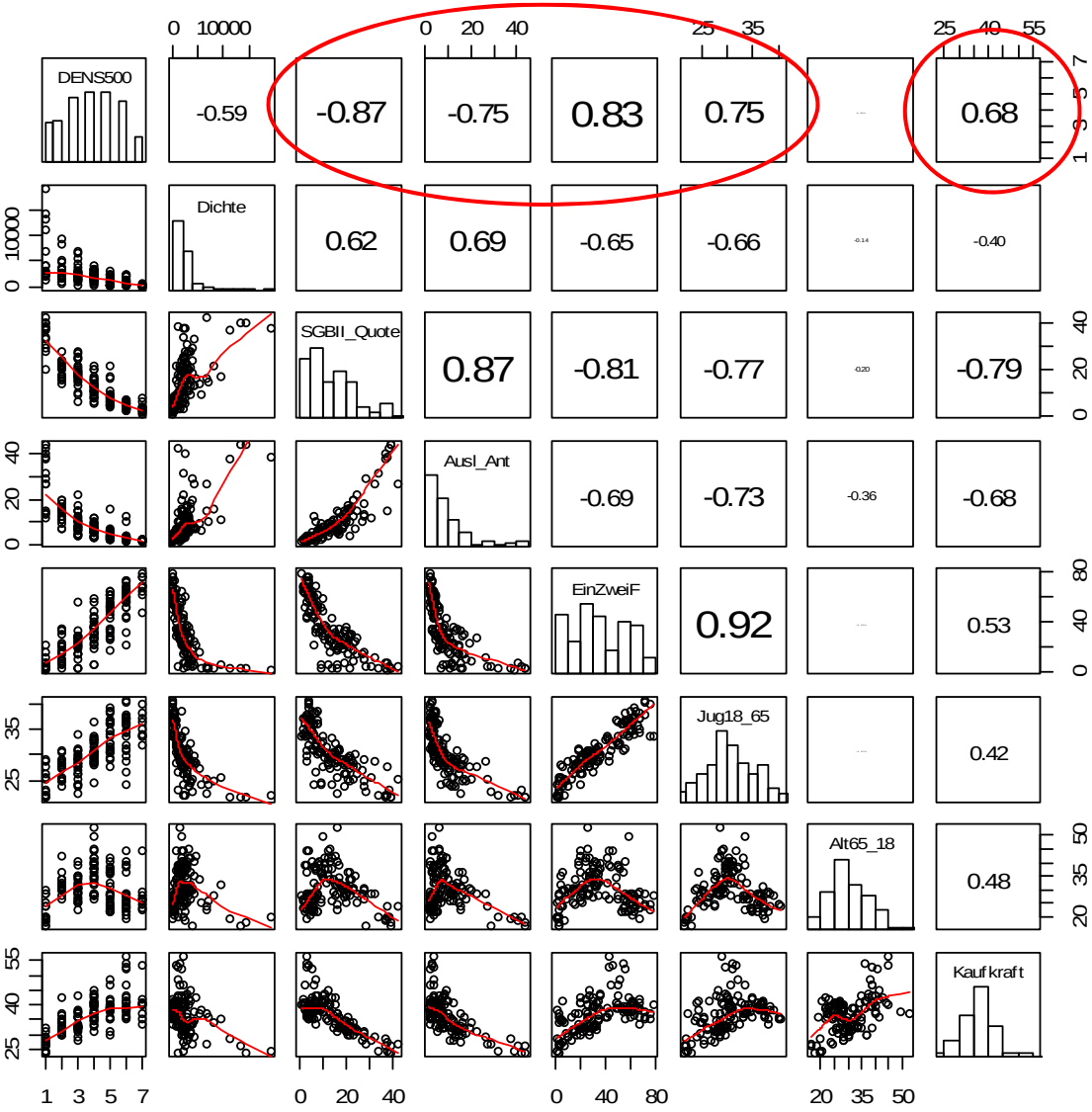
Korrespondenzadresse:

PD Dr. Jörg-Peter Schräpler
Information und Technik NRW (IT.NRW)
Mauerstrasse 51
40476 Düsseldorf

Email: joerg-peter.schraepler@it.nrw.de

Pilotregion Dortmund und Kreis Coesfeld

Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik



Korrelationen mit der KDE-Variable ?

Die KDE-Variable „DENS500“ korreliert hoch mit:

- Nichtdeutschenanteil
- SGBII – Quote
- Anteil Ein-Zweifamilien-Häuser
- Jugendquotient
- Kaufkraft

mittelstark mit der Einwohnerdichte „Dichte“

Rauminformation: Einzugsbezirke der Grundschulen

BMBF Projekt - Erklärte Varianz bei verschiedenen Raumebenen

Erklärte Varianz des IFS - Index auf Basis verschiedener Raumebenen und Schultypen

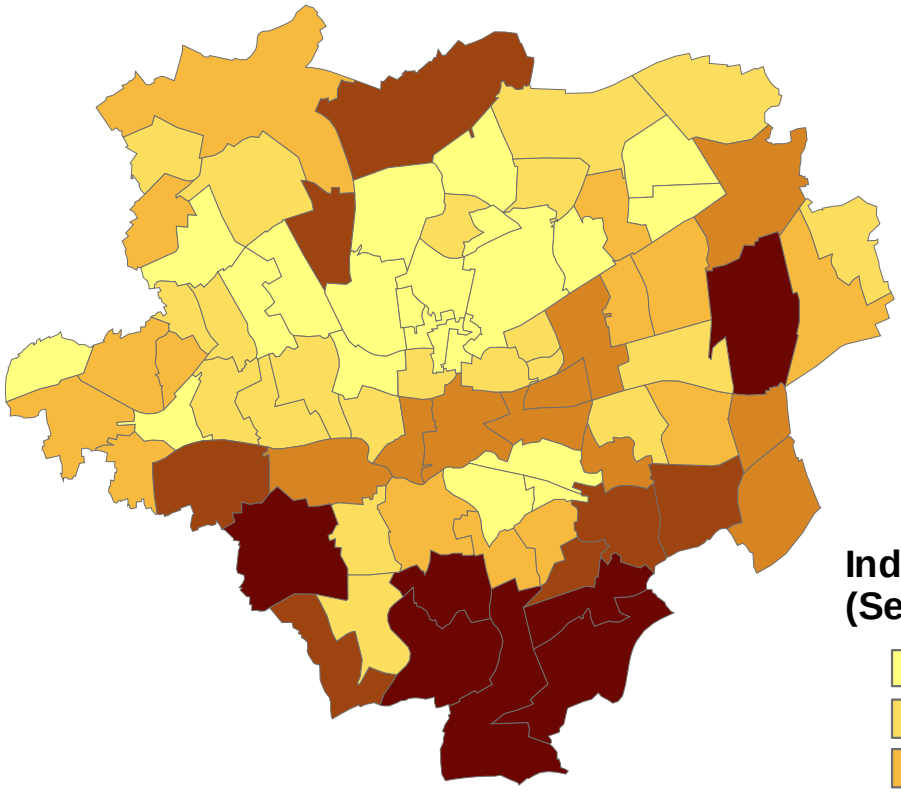
	Grundschulen					
	Raumindex			**Kombinierter Index**		
Raumebene	*Dortmund*	*Coesfeld*	*Insgesamt*	*Dortmund*	*Coesfeld*	*Insgesamt*
echte Einzugsgebiete	0.592	0.011	**0.473**	0.744	0.006	**0.680**
aggregierte Marktzellen (Basis Microm)	0.497	0.017	**0.399**	0.720	0.002	**0.656**
Ortsteile	0.542	0.009	**0.512**	0.712	0.058	**0.683**
Kernel SGB II Dichte	0.567	0.061	**0.514**	0.738	0.108	**0.696**
	weiterführende Schulen					
	Raumindex			**Kombinierter Index**		
Raumebene	*Dortmund*	*Coesfeld*	*Insgesamt*	*Dortmund*	*Coesfeld*	*Insgesamt*
echte Einzugsgebiete	0.269	0.027	**0.263**	0.698	0.686	**0.697**
aggregierte Marktzellen (Basis Microm)	0.217	0.005	**0.239**	0.679	0.673	**0.684**
Ortsteile	0.253	0.034	**0.283**	0.670	0.670	**0.697**
Kernel SGB II Dichte	0.361	0.025	**0.318**	0.724	0.661	**0.717**

Quelle: LDS NRW / ZEFIR / Uni Wuppertal 2008, BMBF - Verbundprojekt

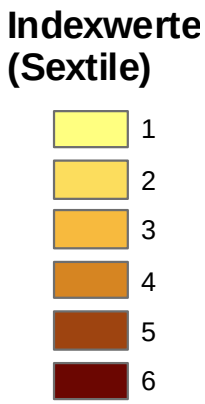
Information und Technik
Nordrhein-Westfalen
Geschäftsbereich Statistik

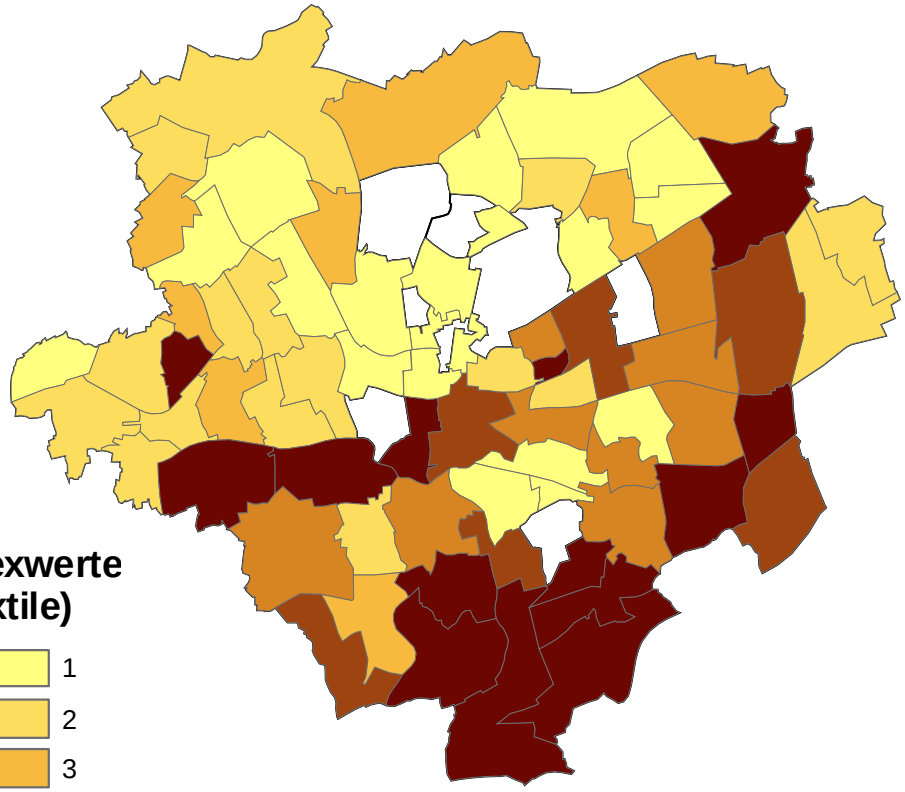
Evaluierung des kombinierten Index anhand des IFS – Index (Sextile)

Dortmunder Gemeinschafts-Grundschulbezirke



Sozialraumindex - Sextile





IFS - Index - Sextile

KDE- Ansatz Wiesbaden, 14.01.2011