

Hot-Spot-Analyse zur Unterstützung der Datenklassifizierung – ein Versuch zur Minimierung der Subjektivität in thematischen Karten am Beispiel des Projekts AlpES (Alpine Ecosystem Services)

PETER FRECKMANN¹

Zusammenfassung: Die Erstellung thematischer Karten setzt in vielen Fällen die Klassifizierung der zu visualisierenden Geodaten voraus. Je nach Wahl einer Klassifizierungsmethode unterscheiden sich die erstellten thematischen Karten mehr oder weniger stark voneinander. Der Kartenersteller muss entscheiden, welche der verschiedenen Klassifizierungsmethoden die Realität am besten abbildet. Die Entscheidung für eine bestimmte Methode impliziert damit eine subjektive Komponente. Der Beitrag zeigt am Beispiel von ausgewählten Indikatoren aus dem Projekt AlpES auf, wie mit Hilfe einer Hot-Spot-Analyse die Datenklassifizierung unterstützt werden kann, um das Ergebnis objektiver zu gestalten.

1 Einleitung

Um Geodaten in thematischen Karten zu visualisieren, ist es häufig notwendig, die Daten zu klassifizieren, um durch quantitative Zusammenfassung räumliche Strukturen erkennbar zu machen. Je nach Wahl einer bestimmten Methode unterscheiden sich die erstellten thematischen Karten mehr oder weniger stark voneinander. Es besteht das Problem, aus verschiedenen Klassifizierungsmethoden die jeweils geeignetste auszuwählen, um die Realität bestmöglich abzubilden. Ein Thema kann also auf unterschiedliche Weise visualisiert und interpretiert werden. Die subjektive Komponente durch den Kartenersteller spielt damit eine nicht zu vernachlässigende Rolle (HRUBY 2016). Der Beitrag soll zeigen, wie mit Hilfe räumlicher, im Gegensatz zu nicht-räumlicher statistischer Verfahren die Datenklassifizierung unterstützt werden kann, um die Subjektivität in thematischen Karten zu minimieren.

2 Nicht-räumliche Datenklassifizierung

Die Klassifizierung der Ausgangsdaten für eine thematische Karte ist notwendig, um Informationen über räumliche Verteilung und Muster visualisieren zu können. Das gilt sowohl für Choroplethenkarten (Variation von Farbton, Helligkeit, Sättigung) als auch für punkt-, linien- oder flächenbezogene Diagrammsignaturen (gestufte Größen). Dabei sollte die Klassenbreite so gewählt werden, dass sie mit der Verteilung der Daten korrespondiert. Da thematische Karten heute fast ausschließlich mit Geoinformationssystemen (GIS) oder entsprechenden Tools erstellt

¹ Hochschule Karlsruhe, Fakultät für Informationsmanagement und Medien, Moltkestr. 30, D-76133 Karlsruhe, E-Mail: peter.freckmann@hs-karlsruhe.de

werden, werden dafür in erster Linie die dort standardmäßig zur Verfügung stehenden Methoden der Klassifizierung verwendet. Dazu gehören insbesondere:

- Äquidistante Klassen – die Klassenbreite ist für jede Klasse gleich groß.
- Natural Breaks – die Klassenbreite richtet sich nach Sprüngen in der Datenreihe, dort wo diese auftreten werden die Klassengrenzen gesetzt. Dadurch werden die Unterschiede in den Werten innerhalb der Klassen minimiert.
- Quantile – allen Klassen ist die gleiche Anzahl an Objekten zugeordnet, d.h. sie haben die gleiche Häufigkeit.

Hinzu kommt die Möglichkeit, Klassengrenzen unabhängig von den vorgegebenen Methoden auch interaktiv durch den Kartenersteller festzulegen. Wie schon häufig in der Fachliteratur diskutiert, führt der Einsatz der genannten Methoden zu unterschiedlichen Ergebnissen in der Kartendarstellung, die dadurch z.T. erheblichen Interpretationsspielraum bieten. (ARMSTRONG et al. 2003; BREWER & PICKLE 2002; MONMONIER 1996; MONMONIER 2005). Allen Methoden gemeinsam ist die Tatsache, dass hier nicht-räumliche statistische Verfahren zur Klassenbildung herangezogen werden (HRUBY 2016). Damit spiegelt der Karteninhalt entweder die Vorstellung des Kartenerstellers über die reale Situation wider, oder es besteht die Gefahr, dass die im GIS standardmäßig voreingestellte Methode übernommen wird, ohne zu hinterfragen, ob diese die Realität am besten abbildet.

3 Räumliche Datenklassifizierung

Im Gegensatz zur nicht-räumlichen Statistik wird in der räumlichen Statistik die subjektive Komponente bei der Bildung von Gruppen reduziert, indem das Erkennen und Zusammenfassen einander ähnlicher und räumlich zusammenhängender Gebiete unterstützt wird (HRUBY 2016). Auch hierzu sei auf die einschlägige Literatur verwiesen (ARMSTRONG et al. 2003; CHEN et al. 2008; DUQUE et al. 2006; KRONENFELD & WONG 2017). Im Folgenden soll ein pragmatischer Ansatz vorgestellt werden, der es erlaubt, mit Hilfe entsprechender GIS-gestützter Funktionen eine Klassenbildung unter Berücksichtigung räumlicher Zusammenhänge durchzuführen (BENNETT et al. 2010).

3.1 Räumliche Autokorrelation – Moran's I Index

Die geographische Zusammenfassung der Objekte (Punkte, Linien, Flächen) eines Untersuchungsgebietes zu Gruppen, erfolgt in zwei Schritten. Zunächst kann mit dem Moran's I Index die räumliche Autokorrelation unter den Objekten bezogen auf die Lage und gleichzeitig auf die Merkmalsausprägung gemessen werden. Der Indexwert weist darauf hin, ob die von den Objekten gebildeten Muster zufällig verteilt oder gruppiert sind (BAHRENBERG et al. 2008; GETIS & ORD 1992; GOODCHILD 1986). Der Moran's I Index basiert auf einer schlussfolgernden Statistik, d.h. die Ergebnisse der Analyse werden immer im Kontext der Null-Hypothese interpretiert. Bei der Moran's I Statistik gibt die Null-Hypothese an, dass das analysierte Attribut zufällig auf die Objekte im Untersuchungsgebiet verteilt wird. Wenn der von der Analyse zurückgegebene p-Wert statistisch signifikant ist, kann die Null-Hypothese abgelehnt werden.

Der Moran's I Indexwert nimmt dabei Werte zwischen -1 und + 1 an. Ein positiver Indexwert zeigt eine Tendenz zur Clusterbildung und ein negativer Indexwert eine Tendenz zur Streuung an. Der Indexwert lässt sich wie folgt berechnen:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \quad (1)$$

mit

z_i	Abweichung des Attributwertes des Objekts i vom Mittelwert
$\omega_{i,j}$	räumliche Gewichtung (Distanz) zwischen den Objekten i und j
n	Anzahl der Objekte
S_0	Summe über alle Distanzen

3.2 Hot-Spot-Analyse – Getis-Ord G_i^* -Statistik

Zeigt der Moran's I Indexwert eine Tendenz zur Gruppierung, lassen sich im zweiten Schritt die Objekte über eine Clusteranalyse zu zusammenhängenden Gebieten aggregieren. Mit Hilfe der Getis-Ord G_i^* -Statistik lässt sich feststellen, wo Objekte mit hohen oder niedrigen Werten zu räumlicher Clusterbildung neigen (GETIS & ORD 1995). Dabei wird jedes einzelne Objekt unter Berücksichtigung der benachbarten Objekte bewertet:

$$G_i^* = \frac{\sum_{j=1}^n \omega_{ij} x_j - X \sum_{j=1}^n \omega_{ij}}{S \sqrt{\frac{[n \sum_{j=1}^n \omega_{ij}^2 - (\sum_{j=1}^n \omega_{ij})^2]}{n-1}}} \quad (2)$$

mit

x_j	Attributwert des Objekts j
$\omega_{i,j}$	räumliche Gewichtung (Distanz) zwischen den Objekten i und j
n	Anzahl der Objekte
X	Arithmetischer Mittelwert des Attributs x
S	Standardabweichung

Ein statistisch signifikanter Hot-Spot ist ein Objekt mit einem hohen Attributwert, das außerdem von anderen Objekten mit hohen Werten umgeben ist. Für die identifizierten Cold-Spots verhält es sich umgekehrt. Das Ergebnis der Analyse zeigt damit gleichzeitig die für die jeweilige Forschungsfrage relevanten Gebiete. Damit ist also auch ein Regionalisierungseffekt verbunden.

4 Anwendungsbeispiel - AlpES

Das beschriebene Verfahren wurde im Interreg Alpine Space Projekt AlpES – Alpine Ecosystem Services getestet (<http://www.alpine-space.eu/projects/alpes/en/home>). Im Rahmen dieses Projekts werden Ökosystemleistungen als Hauptsäulen einer grünen Wirtschaft im Alpenraum untersucht, um Ökosysteme und ihre Leistungen für die Gesellschaft zu schützen und nachhaltig

nutzbar zu machen. Unter Ökosystemleistungen versteht man die Bereitstellung von Gütern und Leistungen durch die Natur. Darunter fallen sowohl Versorgungsleistungen mit Verbrauchsgütern wie etwa Nahrungsmittel oder Holz, als auch die von der Natur bereitgestellten Regulierungsleistungen wie die Reinigung von Luft und Wasser (TAPPEINER & TASSER 2017). Ziel des Projektes ist es, ein Verständnis für Ökosystemleistungen zur Steuerung regionaler und transnationaler Umweltpolitik zu etablieren. An dem Projekt sind 10 Partner aus allen Alpenanrainerstaaten unter der Leitung der EURAC Research (Bozen) beteiligt. Das Gesamtbudget beträgt 2,3 Mio. Euro. Das Untersuchungsgebiet umfasst 16.785 Gemeinden, für die 8 umweltrelevante Indikatorengruppen ausgewertet werden (Abb. 1). Einen wichtigen Teil des Projekts bildet das Work Package „Mapping & Assessment“. Hier werden auf Basis der Indikatoren eine Vielzahl an thematischen Karten erstellt, aus denen Handlungsempfehlungen für die Umweltpolitik abgeleitet und die in einem WebGIS der Öffentlichkeit zugänglich gemacht werden. Beides setzt thematische Karten voraus, die möglichst wenig Spielraum für Fehlinterpretationen bieten sollten. Das gilt sowohl für die Nutzung der Karten durch die Projektbeteiligten als auch für die späteren Nutzer der Projektergebnisse mit unterschiedlich fachlichen Ausrichtungen und regionalen, insbesondere grenzüberschreitenden Kenntnissen.

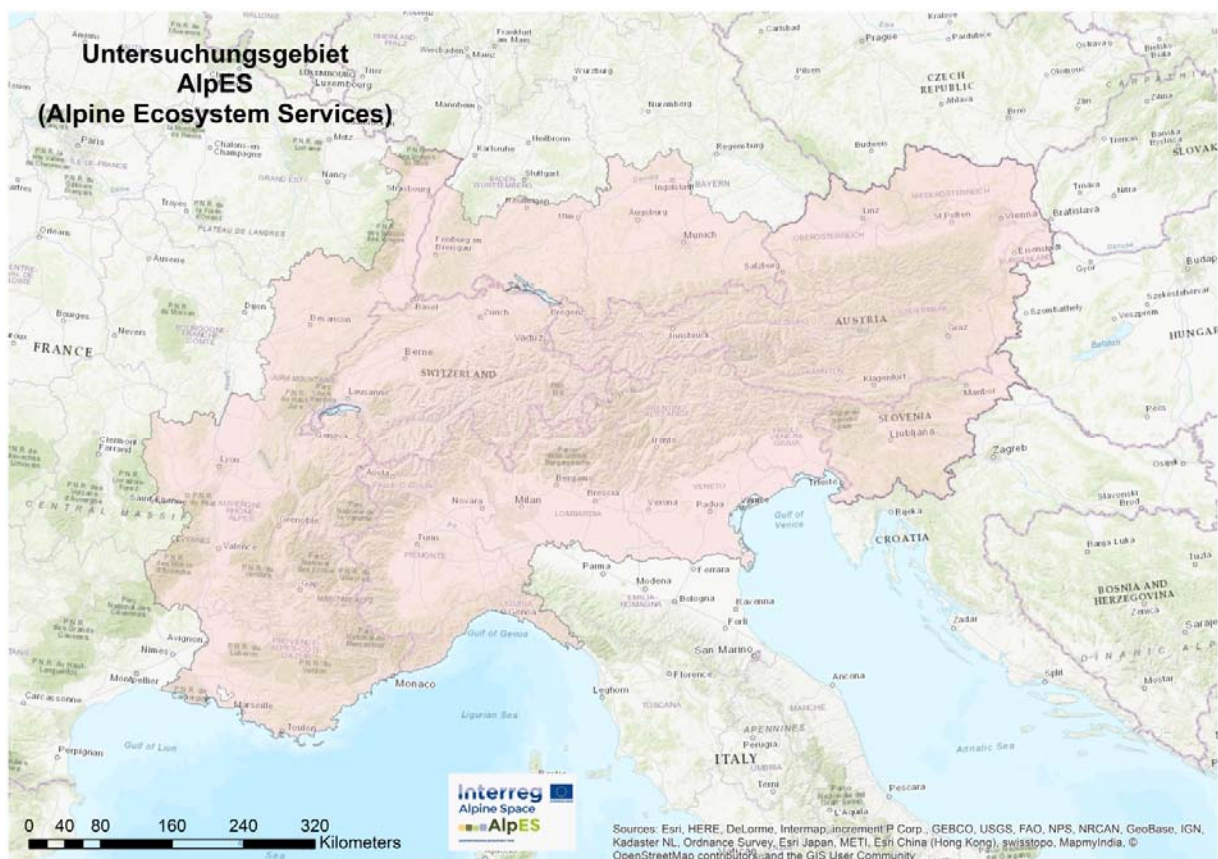


Abb. 1 AlpES Untersuchungsgebiet

4.1 Verfahrensablauf

Abb. 2 verdeutlicht den Ablauf zur Erstellung thematischer Karten. Bei Verwendung nicht-räumlicher Statistik werden die Daten mit den standardmäßigen Methoden auf Klassen verteilt und unter Verwendung kartographischer Modelle in Karten visualisiert. Beim Einsatz räumlicher Statistik wird zunächst der Moran's I Index für die darzustellende Variable bestimmt. Zeigen die Objekte eine Tendenz zur Gruppenbildung wird im nächsten Schritt die Hot-Spot-Analyse durchgeführt. Als Ergebnis erhält man zusammenhängende Flächen mit hohen bzw. niedrigen Attributwerten sowie Gebiete die keine signifikanten Gruppen bilden. Wie später an den Anwendungsbeispielen noch gezeigt wird, lassen sich aus der Analyse Klassengrenzen ableiten, die die Wertespanne der Merkmalsausprägung zwischen den Hot- und Cold-Spots darstellen. Man hat jetzt die Möglichkeit, die standardmäßige Klassenbildung mit dem Ergebnis der Hot-Spot-Analyse zu vergleichen, um auf dieser Grundlage, die Karte mit der am besten geeigneten Klassifizierungsmethode zu verwenden oder aber die Klassen mit den aus der Hot-Spot-Analyse abgeleiteten Klassengrenzen zu definieren. Damit hätte man eine Karte erstellt, deren Datenklassifizierung auf räumlicher Statistik beruht.

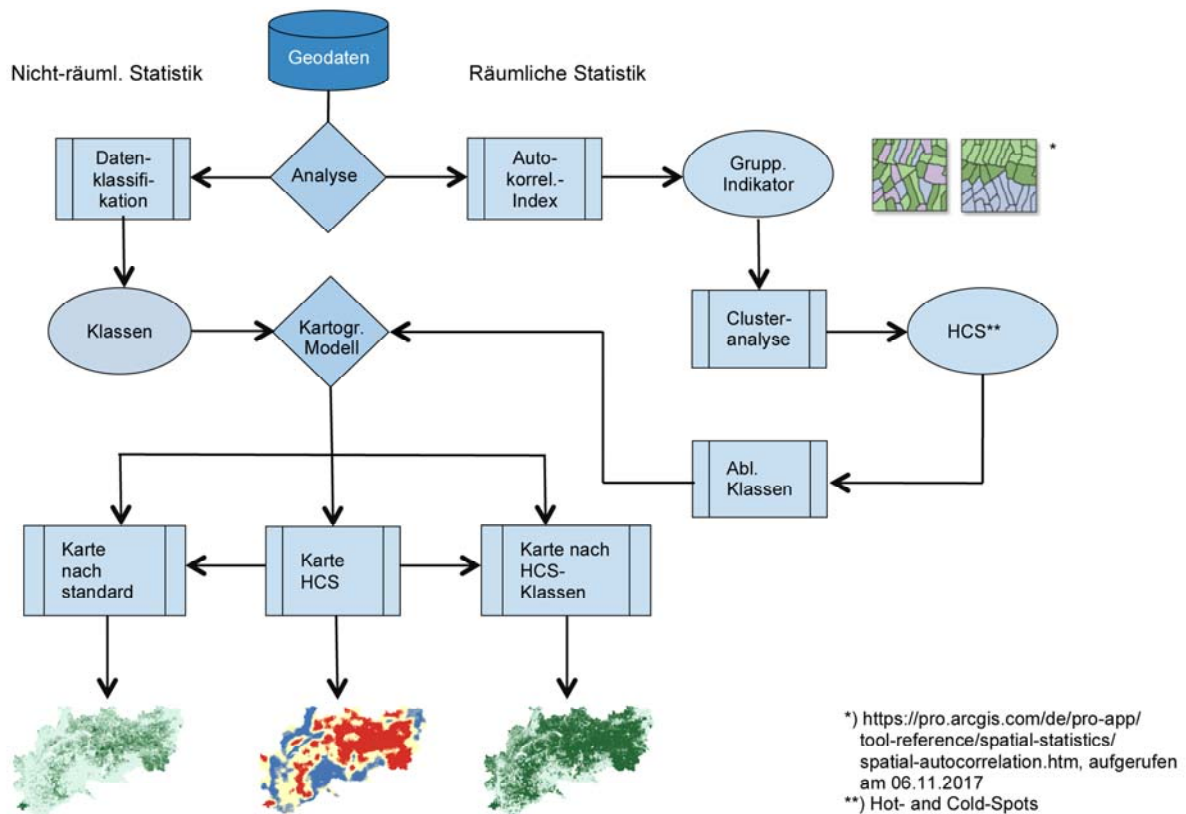


Abb. 2: Erstellung thematischer Karten mit nicht-räumlicher und räumlicher Statistik

4.2 Ergebnisse für ausgewählte Indikatoren

Für die Anwendung der Hot-Spot-Analyse wurden aus den 8 Indikatorengruppen vier Einzel-Indikatoren ausgewählt. Tab. 1 zeigt den für diese Indikatoren jeweils berechneten Moran's I Indexwert. Alle Werte sind positiv und bewegen sich zwischen 0,72 und 0,03. Daraus lässt sich schließen, dass die Gemeindegebiete mehr oder weniger stark zur Gruppenbildung tendieren.

Tab. 1: Indikatoren

AlpES ausgewählte Indikatoren	Moran's I Indexwert
Annual Average Available Water Runoff with Drinking Water Quality	0,72
Population Density (Inhabitants/km ²)	0,55
Fodder Production of Alpine Grassland, Supply of DM per Area	0,03
Grassland (ha)	0,17

Abb. 3 gibt einen Überblick über die Häufigkeitsverteilung der Daten. Dabei wird deutlich, dass sich insbesondere bei den Indikatoren „Population Density“ und „Fodder Production“ der größte Teil der Merkmalsausprägungen im Bereich kleiner Werte liegt und nur wenige Gemeinden sehr hohe Werte aufweisen, wodurch die Klassenbildung mit den Standard-Klassifizierungsmethoden erschwert wird.

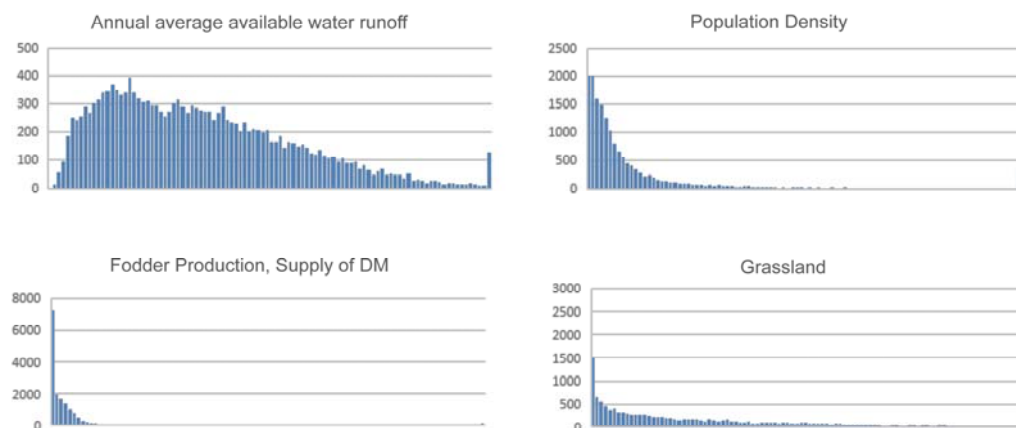


Abb. 3: Verteilung der ausgewählten Indikatoren

Im Folgenden wurden für die vier ausgewählten Indikatoren jeweils 5 thematische Karten erstellt, die das Ergebnis der Hot-Spot-Analyse und verschiedene Klassifizierungsmethoden zeigen. Alle Berechnungen und die Erstellung der Karten wurden mit der ArcGIS Spatial Statistics Toolbox durchgeführt. Vergleichbare Verfahren stehen natürlich auch in anderen Geoinformationssystemen zur Verfügung, so dass das verwendete Tool austauschbar ist.

Abb. 4 zeigt den Mittleren Oberflächenwasserabfluss im Untersuchungsraum. Die Karte der Hot-Spot-Analyse spiegelt das Relief exakt wider. Insbesondere im Bereich der zentralen Alpen liegen die Hot-Spots, da hier der höchste Oberflächenwasserabfluss auftritt. ArcGIS ermittelt die Hot- und Cold-Spots jeweils für das 90, 95 und 99% Signifikanzniveau. Durch die Aggregation der im Untersuchungsraum verteilten Hot- bzw. Cold-Spots für jedes Signifikanzniveau und gleichzeitiger Berechnung des Mittelwertes der jeweiligen Merkmalsausprägung für die aggregierten Flächen, erhält man Mittelwerte, die die Wertespanne der Merkmalsausprägung zwischen den Hot- und Cold-Spots repräsentieren. Diese können als Grundlage für eine Klassenbildung verwendet werden. Damit ergibt sich, wie in Karte b) zu sehen ist, ein fast identisches Bild wie in der Karte der Hot- und Cold-Spots. Die Klassifizierungsmethoden Äquidistante Klassen, Karte c) und Natural Breaks, Karte d) zeigen die räumliche Struktur weniger deutlich. Dagegen entspricht die Karte mit der Methode Quantile sehr genau dem Ergebnis der Hot-Spot-Analyse.

Abb. 5 enthält die Karten für den Indikator Einwohnerdichte. Die auf Basis der Hot-Spot-Analyse festgelegten Klassengrenzen, Karte b) zeigen die auch in Karte a) zu sehenden Kernstädte im Untersuchungsraum und darüber hinaus eine detaillierte Differenzierung der sie umgebenden Agglomerationsräume. Die Klassifizierungsmethoden Äquidistante Klassen und Natural Breaks liefern dagegen ein Ergebnis, dass die reale Situation nicht besonders gut wieder gibt. Quantile dagegen zeigen wiederum eine Situation, die der Realität sehr nahe kommt.

Abb. 6 zeigt die Karten für den Indikator Futterproduktion pro ha. Die Hot-Spot-Analyse ergab nur einige wenige kleine Gebiete mit hohen Werten. Der überwiegende Teil der Gemeinden ist nicht signifikant hinsichtlich einer Gruppierung zu Hot- oder Cold-Spots. Verwendet man die auf Basis der Hot-Spot-Analyse ermittelten Klassengrenzen, ergibt sich ein interessantes Bild. Die Hot-Spots finden sich als einzelne Gemeindeflächen wieder. Hier handelt es sich um Grünlandflächen innerhalb dieser Gemeinden, die sowohl von der Höhenlage als auch den klimatischen Bedingungen besonders gute Bedingungen haben und damit eine hohe Produktivität aufweisen, aber nur vereinzelt auftreten. Die in Deutschland und in der Schweiz zu erkennenden größeren zusammenhängenden Flächen mit höheren Werten, die man in der Karte der Hot-Spot-Analyse nicht sieht, gehören zu größeren Almgebieten, die als Futterflächen mit hoher Energieausbeute ausgewiesen worden sind. Insofern ergibt sich eine realitätsnahe Abbildung. Deutlich wird auch, dass Äquidistante Klassen und Natural Breaks keine brauchbaren Ergebnisse liefern und die Klassifizierung mit Quantilen die Situation visuell überbewertet.

Wurden mit den drei beschriebenen Indikatoren relative Werte betrachtet, soll mit Abb. 7 auch auf Absolutwerte eingegangen werden, da auch Diagramme in gestuften Größen klassifizierte Werte repräsentieren können. Die Karten b) bis e) zeigen einen Ausschnitt des Untersuchungsraums, da die Kreisdiagramme in der Gesamtansicht kaum einzeln sichtbar wären. Zum besseren Vergleich wurden in den Kartenausschnitten die Hot- und Cold-Spots als weitere Ebene unter die Kreisdiagramme gelegt. Deutlich wird hier, dass die Klassifizierung auf Basis

der Hot-Spot-Analyse wieder sehr gut geeignet ist. Dort wo die Hot-Spots liegen, findet man die größten Kreise und umgekehrt. Äquidistante Klassen geben diese Struktur nicht wieder, Natural Breaks nur eingeschränkt. Wohingegen Quantile in etwa der Karte b) entsprechen, also die Situation gut abbilden.

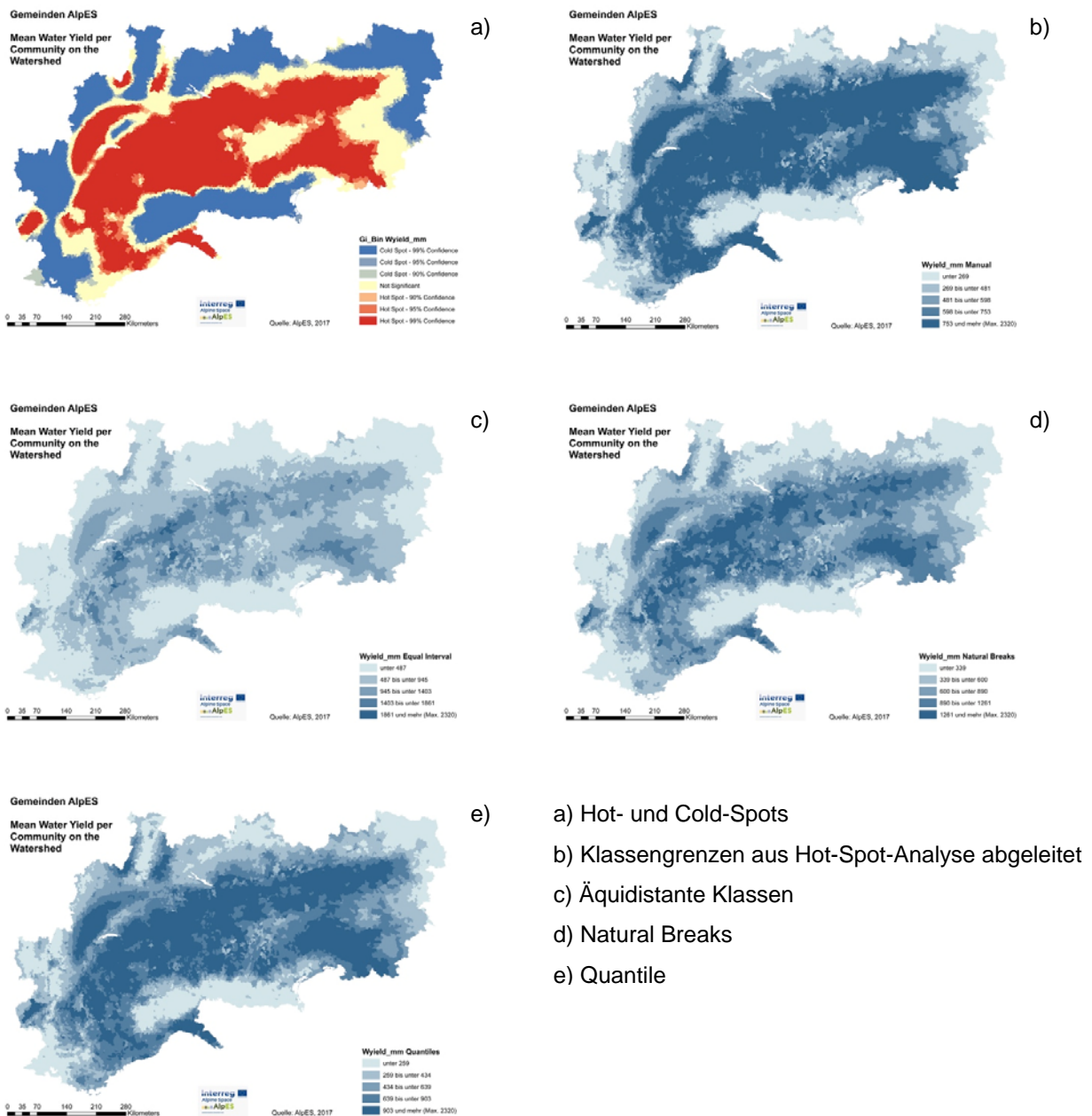


Abb. 4: Thematische Karten für den Indikator Annual Average Available Water Runoff

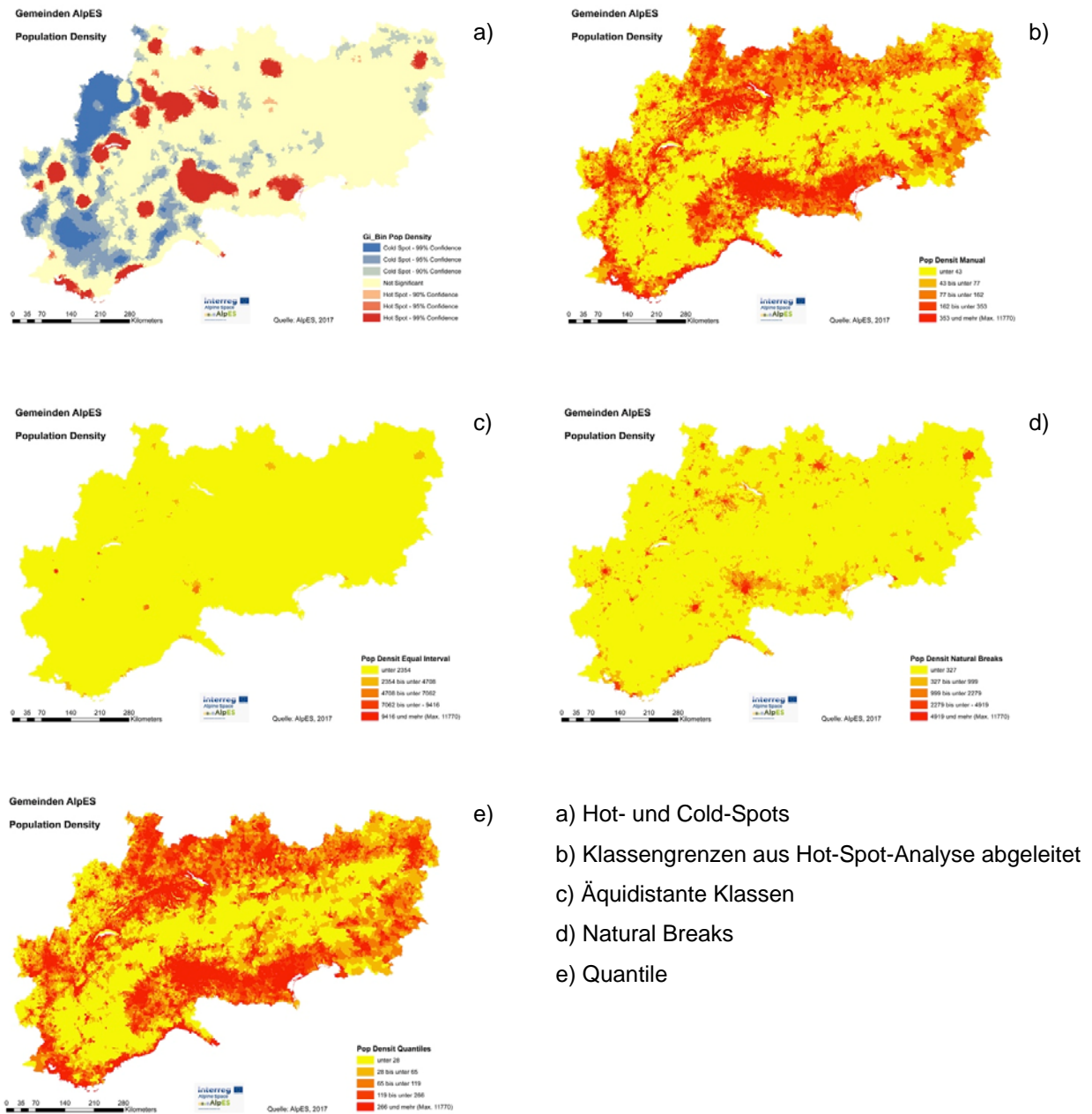
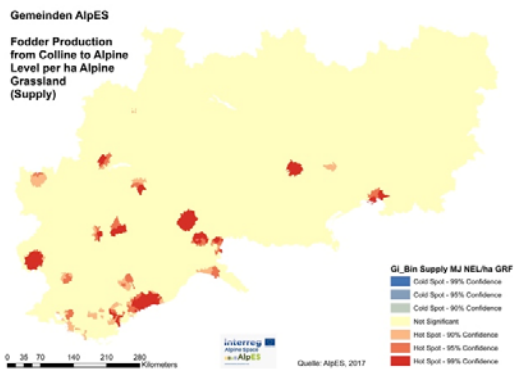
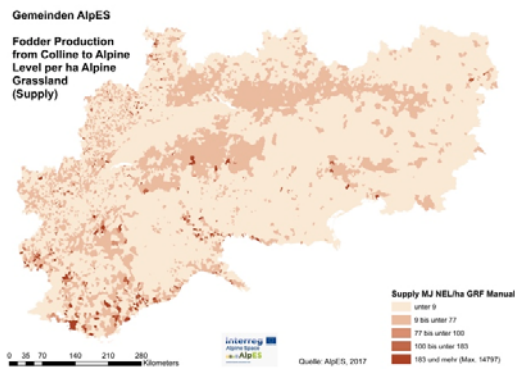


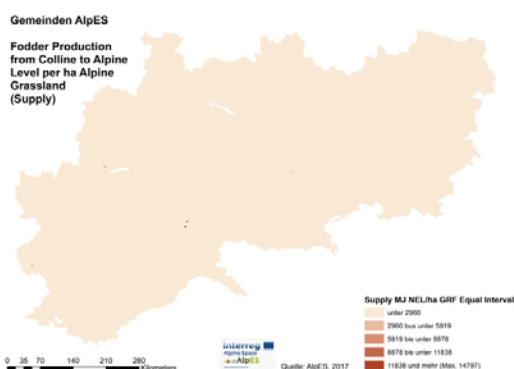
Abb. 5: Thematische Karten für den Indikator Population Density



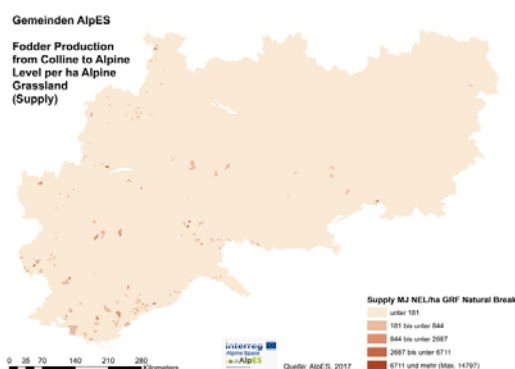
a)



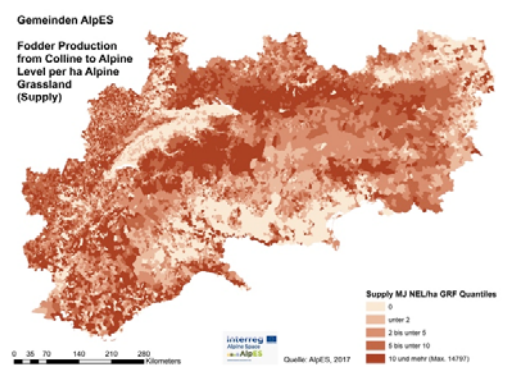
b)



c)



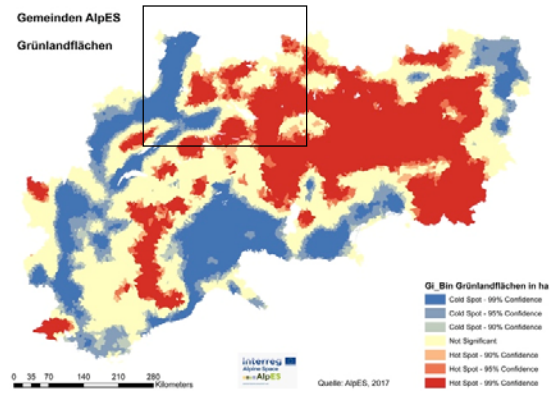
d)



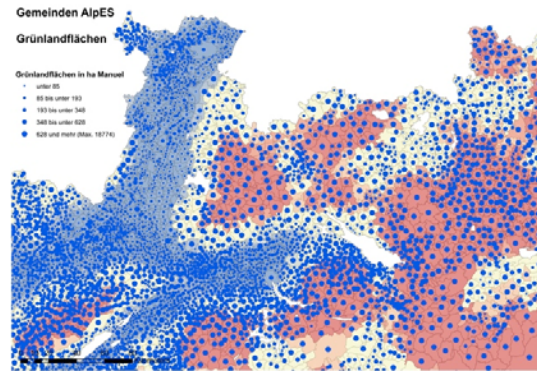
e)

- a) Hot- und Cold-Spots
- b) Klassengrenzen aus Hot-Spot-Analyse abgeleitet
- c) Äquidistante Klassen
- d) Natural Breaks
- e) Quantile

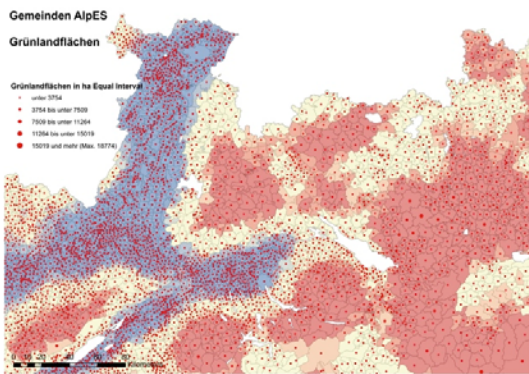
Abb. 6: Thematische Karten für den Indikator Fodder Production of Alpine Grassland



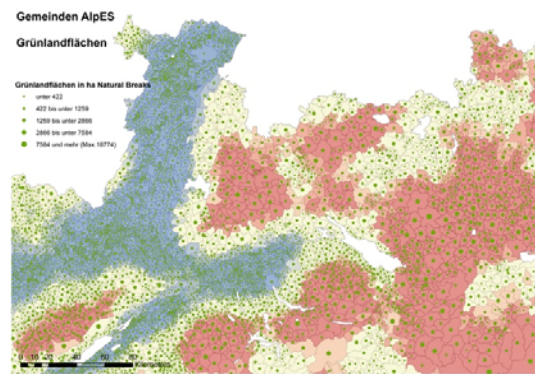
a)



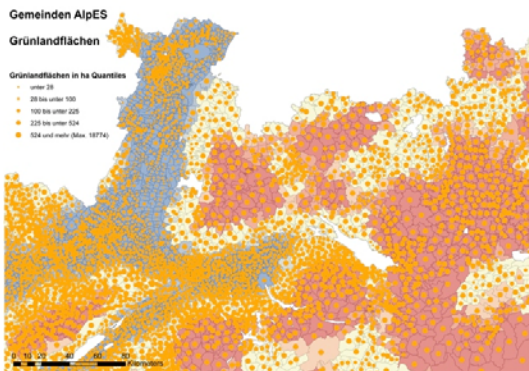
b)



c)



d)



e)

- a) Hot- und Cold-Spots
- b) Klassengrenzen aus Hot-Spot-Analyse abgeleitet
- c) Äquidistante Klassen
- d) Natural Breaks
- e) Quantile

Abb. 7: Thematische Karten für den Indikator Grassland

5 Zusammenfassung

Betrachtet man die oben beschriebenen Ergebnisse, lässt sich bezogen auf die verwendeten Beispieldaten festhalten, dass die Hot-Spot-Analyse eine Klassifizierung mit den Standardmethoden Äquidistante Klassen, Natural Breaks und Quantile dahingehend unterstützt, dass die Subjektivität, die mit den nicht-räumlichen Klassifizierungsmethoden verbunden ist, durch die Gegenüberstellung mit den Ergebnissen der Hot-Spot-Analyse, reduziert werden kann. Im Einzelnen lassen sich folgende Vorteile erkennen:

- Mit der Hot-Spot-Analyse lässt sich, wie generell mit Clusterverfahren, eine Regionalisierung innerhalb des Untersuchungsraumes durchführen. Hot- und Cold-Spots bilden bezogen auf ein bestimmtes Merkmal weitgehend homogene Gebiete. Das ist umso ausgeprägter, je höher der Moran's I Index ausfällt.
- Bei vorliegender Autokorrelation erlaubt es die Hot-Spot-Analyse die standardmäßig zur Verfügung stehenden Klassifizierungsmethoden dahingehend zu überprüfen, ob diese die vorhandenen räumlichen Verteilungen und Muster wieder geben und hilft dabei eine fundiertere Entscheidung für eine der Standardmethoden zu treffen.
- Unabhängig davon besteht auch die Möglichkeit, Klassengrenzen auf Basis der Hot-Spot-Analyse zu bestimmen. Wie die Anwendungsbeispiele gezeigt haben, ergibt sich damit, vorallem bei einer deutlichen Tendenz zur Gruppenbildung der Objekte, ein Ergebnis, das den tatsächlichen realen räumlichen Strukturen näher kommt als insbesondere mit den Standardmethoden Äquidistante Klassen und Natural Breaks.

6 Literaturverzeichnis

- ARMSTRONG, M.P., XIAO, N. & BENNETT, D.A., 2003: Using Genetic Algorithms to Create Multicriteria Class Intervals for Choropleth Maps. *Annals of the Association of American Geographers*, **93**(3), 595-623.
- BAHRENBERG, G., GIESE, E., MEVENKAMP, N. & NIPPER, J., 2008: Statistische Methoden in der Geographie. *Multivariate Statistik*, **2**, 352-370.
- BENNETT, L., D'ACOSTA, J. & VALE, F., 2010: Spatial Data Mining: A Deep Dive into Cluster Analysis. <https://blogs.esri.com/esri/arcgis/2010/07/13/spatial-statistics-resources/>, letzter Zugriff 25.09.2017.
- BREWER, C. & PICKLE, L., 2002: Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. *Annals of the Association of American Geographers*, **92**(4), 662-681.
- CHEN, J., ROTH, R.E., NAITO, A.T., LINGERICH, E.J. & MACEACHREN, A.M., 2008: Geovisual analytics to enhance spatial scan statistics interpretation; an analysis of U.S. cervical cancer morality. *International Journal of Health Geographics*, **7**(57), 1-18
- DUQUE, J.C., RAMOS, R. & SURINACH, J., 2007: Supervised Regionalization Methods: A Survey. *Research Institute of Applied Economics 2006, Working Papers 2006/8*, 1-31.

- GETIS, A. & ORD J.K., 1992: The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* **24**(3), 189-206.
- GETIS, A. & ORD J.K., 1995: Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* **27**(4), 286-306.
- GOODCHILD, M.F., 1986: Spatial Autocorrelation. *Catmog* 47, Geo Books.
- HRUBY, F. 2016: 190 Jahre Choroplethenkarten – Ein Zwischenresümee. *Kartographische Nachrichten* **2/16**, 58-65
- KRONENFELD, B.J. & WONG, D.W.S., 2017: Visualizing statistical significance of disease clusters using cartograms. *International Journal of Health Geographics*, **16**(19), 1-13
- MONMONIER, M. 1996: *How to Lie with Maps*, Second Edition.
- MONMONIER, M. 2005: Lying with Maps, *Statistical Science*, **20**(3), 215-222.
- TAPPEINER, U. & TASSER, E., 2017: Ökosystemleistungen in den Alpen. *Natur und Landschaft*. DOI: 10.17433/9.2017.50153509.440-445.