

Integration von Daten der Umweltbeobachtung mit Metadaten, multivariat-statistischer Naturraumgliederung, Geostatistik und GIS

Winfried Schröder

1 Hintergrund und Zusammenfassung

Die ökologische Umweltbeobachtung soll den Zustand von Ökosystemen in räumlicher und zeitlicher Differenzierung quantifizieren. Voraussetzungen hierfür sind: 1. Räumlich separat erhobene Daten über ökologisch zusammengehörende Sachverhalte müssen fachlich und technisch integrierbar sein. 2. Räumliche und zeitliche Messwertunterschiede müssen den tatsächlichen Gegebenheiten entsprechen und dürfen nicht auf unterschiedliche Datenerhebungsmethoden zurückzuführen sein.

Die Harmonisierung der Umweltbeobachtung erfordert also neben der Standardisierung der empirischen Methoden vor allem die Berücksichtigung der räumlichen Repräsentanz von Messdaten und Messnetzen. Deshalb ist es zunächst zweckmäßig, Metadaten über vorhandene Umweltmessnetze zu erheben und auszuwerten, die zur Beantwortung folgender Fragen erforderlich sind: Wo wird gemessen? Was wird gemessen? Wie wird gemessen? Entsprechende Informationen über Umweltmessnetze des Bundes und der Länder wurden mit einem rund 800 items umfassenden elektronischen Fragen erhoben und die hiermit aufgebaute Metadatenbank an ein GIS gekoppelt. Das GIS enthält ökologische Flächendaten über Klima, Boden und die potenziell natürliche Vegetation sowie eine daraus multivariat-statistisch abgeleitete Raumgliederung. Die Raumgliederung dient der Bestimmung der häufigkeitsstatistischen und raumstrukturellen Landschaftsrepräsentanz der Messnetze. Diese *Messnetzrepräsentanz* wird ergänzt durch die geostatistische Prüfung der *Messdatenrepräsentanz*. Die kombinierte

Anwendung dieser Instrumente wird am Beispiel von Fragestellungen aus der Praxis anhand von Messdaten aus der Umweltbeobachtung gezeigt.

2 Metadatenerhebung

Die Metadaten zu den Umweltmessnetzen wurden mit einem rund 800 Items umfassenden elektronischen Fragebogen folgender Struktur erhoben: (1) Allgemeine Angaben, (2) Boden, (3) Luft, (4) Wasser, (5) Bioindikation, (6) Umwelt und Gesundheit, (7) Landschaft und Natur sowie (8) Gentechnik. Die Metadaten zu Bodendauerbeobachtungsflächen (Zf. 2), zum Monitoring der Schwermetallakkumulation in Moosen (= ein Messnetz unter Zf. 5) sowie zur Uurden von allen Bundesländern geliefert. Darüber hinaus lieferten Baden-Württemberg, Bayern, Brandenburg, Hamburg, Hessen, Saarland und Schleswig-Holstein Informationen zu den weiteren item-Kategorien des elektronischen Fragebogens. Diese Metadaten bilden das derzeitige Fundament für die Identifikation von inhaltlichen und geographischen Lücken in den deutschen Umweltmessnetzen, Repräsentanzuntersuchungen und fragestellungsbezogene Verknüpfungen der Meta- und Messdaten der Monitoringnetze mit landschaftsökologischen Flächendaten.

Für jede Fragebogendatei musste jeweils eine spezifische Verfahrensweise zur Anbindung der Metainformation an das jeweilige Messnetz entwickelt und durchgeführt werden. Für die Anbindung der Metadaten in ArcView, die in der Tabelle "ArcViewInfos" in MS Access abgelegt sind, ist es erforderlich, die Daten in einer externen Datenbank (Datenaustauschformat dBase IV) zu speichern. Der daran anschließende Verfahrensschritt besteht lediglich im Import dieser Datenbank in ArcView. Nach dem Import der Metadaten werden diese mit den Standorten der einzelnen Messnetze unter Verwendung des eigens für diese Zwecke programmierten "Linking-Tools" zusammengeführt: Alle Metainformationen über die Messnetze werden mit den entsprechenden Messstandorten verknüpft. Dies erleichtert dem Anwender die übersichtliche Zusammenstellung aller Informationen über die Messnetze. Ein Klick auf einen Standort eines Messnetzes bringt alle Fragen und Antworten zum Vorschein, die für diesen Punkt im Fragebogen formuliert wurden. Anhand der Fragenummer in der Informationsbox in ArcView lässt sich die Zusammengehörigkeit mit dem Fragebogen erkennen. Für alle Messnetze und

Metadaten wurden dementsprechend ArcView-Projekte unter Verwendung einer einheitlichen Benennungskonvention erzeugt. Um die Messnetzmetadatenbanken umfassend und übergreifend auf ihre Inhalte untersuchen zu können, wurden drei Abfragewerkzeuge entwickelt: 1. Analyse von Messnetzen in Bezug auf Messgrößen und Messmethoden. 2. Such- und Informationsdialog (MeSID) für die räumliche und inhaltliche Abfrage der Metadaten unter Verwendung von Schlagworten im GIS. Damit kann beispielsweise ermittelt werden, welche Messgröße wo und mit welcher Methode erfasst wird. 3. Abgleich der erhobenen Metadaten mit dem Kerndatensatz nach Bosch und Partner (2002) sowie Haber et al. (1997), der diejenigen Messgrößen beinhaltet, die durch die Umweltbeobachtung erhoben werden sollten. Damit lassen sich für ausgewählte Fragen des Kerndatensatzes die Entsprechungen des Fragebogens Umweltbeobachtung herausfiltern sowie diejenigen Ländermessnetze anzeigen, die Messgrößen dieser Schnittmenge erheben.

3 Landschaftsökologische Raumgliederung

Anforderung. Die Messnetze der Umweltbeobachtung sollen die wichtigsten landschaftsökologischen Raumeinheiten abdecken (Landschaftsrepräsentanz). Zur statistischen Überprüfung dieses Kriteriums wird eine landschaftsökologische Raumgliederung benötigt.

Daten und statistisches Verfahren. Die Raumgliederung wurde mit Classification and Regression Trees (*CART*) berechnet. Das *CART*-Modell beschreibt die räumliche Differenzierung der in 67 Kategorien gegliederten potenziell natürlichen Vegetation mit Flächendaten über die räumliche Struktur folgender Landschaftsmerkmale: Höhenlage, Klima (je 12 über das Zeitintervall 1961 – 1990 integrierende Monatsmittelwerte von Lufttemperatur, Niederschlag, Verdunstung; 9 über das Intervall 1981 – 1989 integrierende Mittelwerte der Globalstrahlung) und Bodenart (25 Kategorien) statistisch beschrieben.

CART ermöglicht es, sehr große Mengen kategorialer, ordinaler und metrisch-kontinuierlicher Daten ohne Veränderung ihrer Skalendignität zu verarbeiten. Dies ist wichtig, denn die Raumgliederung Deutschlands wird für rund 88.400 Zellen à 2 x 2 km berechnet, wobei jede von ihnen durch die Ausprägung jeder der insgesamt 48 Landschaftsmerkmale gekennzeichnet ist. Ferner werden von *CART* automatisch alle

mit einem gegebenen Datensatz möglichen Gliederungen ermittelt. Sie sind anhand statistischer Maßzahlen miteinander vergleichbar, und auf dieser Grundlage wird die statistisch optimale Gliederungsvariante ermittelt. Diese ist vom Anwender nach inhaltlichen Kriterien modifizierbar (s.u.). *CART* ist ein binär divisives hierarchisches Verfahren, das zunächst die nach Datenlage maximal möglichen Klassenanzahl ermittelt und in einem Strukturbaum (T_{\max}) veranschaulicht. Anschließend wird durch sukzessives Zurückschneiden (pruning) einzelner Blätter eine Sequenz aus allen möglichen Bäumen T' mit abnehmender Klassenanzahl (Größe des Baumes) und damit zunehmender Fehlklassifikationsrate der T' , T'' usw. bis zur Ausgangsklasse (Grundgesamtheit) erstellt. Für jeden dieser Bäume werden die Fehlklassifikationsraten des gesamten Baums berechnet. *CART* schlägt einen dieser Bäume als statistisch optimalen vor. Dabei wird die Fehlklassifikationsrate mit der Größe des Baums verglichen und daraus ein Komplexitätswert ermittelt (cost complexity). Es wird derjenige Baum als optimal vorgeschlagen, der beide Kriterien minimiert. T_{\max} umfasst 73 Raumklassen, der zur besseren Übersichtlichkeit auf die oberen Gliederungstufen des T_{\max} mit 21 Endpunkten (Blättern) beschränkt. Dies sind die mit einer Code-Nummer bezeichneten Raumklassen, deren geographische Lage als kartografisch abgebildet werden kann.

Landschaftsrepräsentanz. Durch Verschneidung der in der Metadatenbank archivierten Lagekoordinaten der Umweltmessnetze mit der landschaftsökologischen Raumgliederung wird die Belegung der Raumklassen mit Messnetze anhand der Flächenproportionalität quantifiziert (*häufigkeitsstatistische Landschaftsrepräsentanz*). Die hierzu berechneten Ergebnisse verdeutlichen einen unterschiedlich ausgeprägten Bedarf an Messnetzoptimierung. Hierfür werden weitere Verfahren benötigt: die Geostatistik zur Prüfung der Messwertrepräsentanz (Kap. 5) und die Nachbarschaftsanalyse zur Bestimmung der raumstrukturellen Landschaftsrepräsentanz.

Das Ergebnis der häufigkeitsstatistischen Landschaftsrepräsentanzanalyse besagt nichts über das räumliche Muster, in dem Raumausschnitte und ihre beliebig kleinen Teilflächen angeordnet sind. Die Untersuchung von Karten räumlich differenzierter Sachverhalte mit der Nachbarschaftsanalyse ergibt eine quantitative Beschreibung der räumlichen Umgebungsstruktur von Kartenausschnitten in Bezug auf bestimmte Merkmale. Auf dieser Grundlage lassen sich Teilflächen identifizieren, deren

Umgebung hinsichtlich Merkmalsausstattung *und* – anordnung typisch ist (*raumstrukturelle Landschaftsrepräsentanz*). Zu diesem Zweck wird für Teilflächen *einer* Karte ein Repräsentanzindex (*RI*) und für Ausschnitte *mehrerer* Karten ein multidimensionaler Nachbarschaftsrepräsentanzindex (*MNR*) berechnet. *RI* bzw. *MNR* quantifizieren also die raumstrukturelle Repräsentanz von jeder einzelnen der rund $88.400 \text{ 2 x 2 km}^2$ großen Teilflächen jeweils einer von *n* Karten, auf denen die räumliche Differenzierung je eines landschaftsökologischen Merkmals abgebildet ist.

Der *RI / MNR* ist eine dimensionslose Zahl, die ausdrückt, ob die räumliche Struktur der Umgebung eines durch ein Merkmal(bündel) gekennzeichneten Raumausschnitts der durchschnittlichen Nachbarschaft aller durch dasselbe Merkmal(bündel) gekennzeichneten Teilflächen völlig (*RI / MNR* = 1) oder überhaupt nicht entspricht (*RI / MNR* = 0). Damit es ist möglich, 1. die raumstrukturelle Landschaftsrepräsentanz bestehender Messnetzstandorte zu quantifizieren, 2. repräsentative Standorte zur Ergänzung bestehender oder die Einrichtung neuer Umweltmessnetze zu bestimmen und 3. bestehende Messnetze nach dem Kriterium der raumstrukturellen Messstellenrepräsentanz auszudünnen (Kap. 4).

4 Messnetzoptimierung nach geostatistischer Messwertrepräsentanz und raumstruktureller Landschaftsrepräsentanz

Konzept. Umweltbeobachtungsnetze, deren Messdaten geostatistisch valide zwischen den Erhebungspunkten interpoliert werden können, lassen sich anhand der Befunde über ihre häufigkeitsstatistische und raumstrukturelle Landschaftsrepräsentanz so weit ausdünnen und damit effizienter gestalten, dass eine ausreichende Zuverlässigkeit der Flächenschätzungen gewährleistet bleibt. Die *Messwertrepräsentanz* ist das Hauptkriterium für die Bewertung bereits bestehender Messnetze, denn sie betrifft folgende, in Frageform gefassten Aspekte: 1. Lassen sich die an den Messpunkten erhobenen Daten auf Flächen, an denen keine Messwerte erhoben werden, übertragen? 2. Kann die Zuverlässigkeit dieser Schätzwerte quantifiziert werden?

Verfahren. Die Auswertung von Messdaten mit den geostatistischen Verfahren Variogramm-Analyse und Kriging-Interpolation hilft, diese beiden Fragen zu beantworten. Mit der Variogrammanalyse wird die räumliche Autokorrelation der

Messdaten geprüft. Ist diese gegeben, können die Messdaten innerhalb dieses Autokorrelationsraumes (range) flächenhaft mit Kriging-Interpolation verallgemeinert werden. Für jeden interpolierten Wert lässt sich die Schätzvarianz als Maß für die Zuverlässigkeit der Schätzung berechnen. Mit diesem Verfahrensgang werden anschließend die Messwerte der Umweltbeobachtungsmessnetze

Anwendungsbeispiel. Die SO₂-Immission wird von den Bundesländern an insgesamt 480 Standorten untersucht. Für die Umweltberichterstattung des Bundes stellt sich die Frage, mit wie vielen und welchen zusätzlichen Standorten der Bundesländermessnetze das im Jahr 1996 31 Stationen umfassende Immissionsmessnetz des UBA ergänzt werden muss, um geostatistisch valide Aussagen zur Schwefeldioxid-Belastung für das Gesamtgebiet der BRD treffen zu können. Hierfür bietet es sich an, die Auswahl der Ergänzungsstationen anhand ihrer raumstrukturellen Repräsentanz für die standortökologischen Raumeinheiten Deutschlands vorzunehmen. Ergänzt man die 31 UBA-Stationen um die gemäß Nachbarschaftsanalyse repräsentativsten 113 Länderstationen, so zeigt das entsprechende Variogramm eine distanzabhängige räumliche Autokorrelation, die eine anschließende Kriging-Interpolation gestattet. Deren Schätzqualität unterscheidet sich nicht nennenswert von derjenigen Flächenschätzung, die auf der Grundlage der 480 Stationen berechnet wurde.

5 Fazit

Die Geostatistik ermöglicht in Kombination mit der Raumgliederung die Analyse und Bewertung von Umweltmessnetzen im Hinblick auf deren Suffizienz und Effizienz. Zudem stellen geostatistische Flächenschätzungen die methodische Grundlage für die Verknüpfung von Messdaten inkongruenter Umweltbeobachtungsnetze dar.

Literatur

Schröder, W.; Schmidt, G.; Pesch, R. (2003): Harmonization of environmental monitoring. Tools for examination of methodical comparability and spatial representativity. In: Gate to Environmental and Health Sciences, July 2003, pp. 1 – 13 [DOI:<http://dx.doi.org/10.1065/ehs2003.07.010>]