

TEXTE

67/2024

Evaluierung und Weiterentwicklung der Regionalisierungsverfahren zur Ausweisung nitratbelasteter Gebiete nach AVV GeA in Deutschland

Sachverständigengutachten

von:

Prof. Dr. Alexander Brenning

Herausgeber:

Umweltbundesamt

TEXTE 67/2024

Projektnummer 179139

FB001435

Evaluierung und Weiterentwicklung der Regionalisierungsverfahren zur Ausweisung nitratbelasteter Gebiete nach AVV GeA in Deutschland

Sachverständigengutachten

von

Prof. Dr. Alexander Brenning

Im Auftrag des Umweltbundesamtes

Impressum

Herausgeber

Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
Internet: www.umweltbundesamt.de

Durchführung der Studie:

Prof. Dr. Alexander Brenning, c/o Friedrich-Schiller-Universität Jena
Löbdergraben 32
07743 Jena

Abschlussdatum:

Januar 2024

Redaktion:

Fachgebiet II 2.1 Übergreifende Angelegenheiten Wasser und Boden
Falk Hilliges

Publikationen als pdf:

<https://www.umweltbundesamt.de/publikationen>

ISSN 1862-4804

Dessau-Roßlau, April 2024

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen*Autoren.

Vorbemerkung des Herausgebers

Gemäß EU-Nitratrichtlinie (91/676/EWG) sind die Mitgliedsstaaten der Europäischen Union verpflichtet, die Verunreinigung von Gewässern mit Nitrat aus der Landwirtschaft vor allem durch Düngung zu verhindern und vorzubeugen. Als Folge des Urteils des Europäischen Gerichtshofs gegen die Bundesrepublik Deutschland wegen mangelhafter Umsetzung der EU-Nitratrichtlinie (Rechtssache C-543/16) hatte die Bundesregierung im Jahr 2020 die Düngeverordnung novelliert und darin die Ausweisung sogenannter nitratbelasteter Gebiete vorgeschrieben. Dabei handelt es sich um Flächen, auf denen hohe Stickstoffeinträge aus landwirtschaftlicher Bewirtschaftung zu einer erhöhten Konzentration des Grundwassers mit Nitrat führen und in denen verschärfte Regelungen zur Ausbringung von Düngemitteln für Landwirte verpflichtend sind. Mit der Ausweisung dieser Gebiete sollen besonders belastete Flächen identifiziert und durch die reduzierte Düngung die Nitratreinträge in das Grundwasser vermindert werden. Verantwortlich für die Ausweisung der Gebiete sind die Bundesländer.

Die technischen Vorgaben auf welche Art und Weise die Ausweisung von nitratbelasteten Gebieten zu erfolgen hat, ist in einer eigenen, bundesweit gültigen Verwaltungsvorschrift ausgeführt (Allgemeine Verwaltungsvorschrift zur Ausweisung von mit Nitrat belasteten und eutrophierten Gebieten: AVV Gebietsausweisung - AVV GeA). Diese fordert bis Ende 2028 ein bundeseinheitliches geostatistisches Regionalisierungsverfahren anzuwenden. Damit wird der Forderung der EU-Kommission Rechnung getragen, eine bundesweit einheitliche und damit zwischen den Bundesländern vergleichbare Ausweisungsmethodik einzuführen. Übergangsweise gestattet die AVV GeA den Bundesländern bis Ende 2028 unterschiedliche Methodiken zur Ausweisung nitratbelasteter Gebiete anzuwenden. Damit wurde den Bundesländern die Möglichkeit eingeräumt, ihre in der Vergangenheit eingesetzten Regionalisierungsverfahren weiter rechtssicher anzuwenden, diese aber perspektivisch auf ein geostatistisches Regionalisierungsverfahren umzustellen. Neben der Anwendung geostatistischer Verfahren wurden von den Bundesländern in der Vergangenheit auch deterministische Regionalisierungsverfahren (Inverse Distanzgewichtung – IDW, Voronoi-Verfahren bzw. Nächster-Nachbar-Interpolation, NNI) sowie eine Abgrenzung nach hydrogeologischen und/oder hydraulischen Kriterien für Zwecke der Regionalisierung eingesetzt. Die Auswahl und der Einsatz dieser Verfahren erfolgte in den Bundesländern nach dem zum individuellen Zeitpunkt der Einführung vorliegenden fachlichen Wissensstand und den zu diesen Zeitpunkten vorhandenen technischen Möglichkeiten. In einigen Ländern wurde durch wissenschaftliche Gutachten die Verwendung bestimmter Methoden empfohlen.

Das vorliegende Gutachten hat zum Ziel, die Einführung eines bundesweit einheitlichen geostatistischen Regionalisierungsverfahrens vorzubereiten. Dafür ist eine Bestandsaufnahme der bisher verwendeten und nach AVV GeA gültigen Regionalisierungsverfahren notwendig, die die Verfahren hinsichtlich ihrer Eignung zur problemadäquaten Abgrenzung nitratbelasteter Gebiete bewertet sowie Weiterentwicklungsmöglichkeiten und alternative Methoden der räumlichen Statistik identifiziert, die bis Ende 2028 bundesweit einheitlich umgesetzt werden könnten.

Kurzbeschreibung: Evaluierung und Weiterentwicklung der Regionalisierungsverfahren zur Ausweisung nitratbelasteter Gebiete nach AVV GeA in Deutschland

Die Ausweisung nitratbelasteter Gebiete für den Grundwasserschutz auf Grundlage von EU-Nitratrichtlinie (91/676/EWG), Düngeverordnung und AVV GeA sieht die Nutzung geostatistischer oder deterministischer Regionalisierungsverfahren vor. Die aktuelle Ausweisungspraxis auf Landesebene soll in ein bundesweit einheitliches geostatistisches Verfahren überführt werden. Vor diesem Hintergrund zielt dieses Gutachten darauf ab, die bisherigen Verfahren hinsichtlich ihrer Eignung und Weiterentwicklungsmöglichkeiten zu bewerten und zu untersuchen, inwieweit neue oder alternative Methoden der räumlichen Statistik besser geeignet sind und bis Ende 2028 umgesetzt werden könnten.

Eine eingehende Problemanalyse und empirische sowie theoretische Methodenbewertung identifiziert Mängel und mögliche Verzerrungen der bislang verwendeten Verfahren der Inversen Distanzgewichtung (IDW), Nächster-Nachbar-Interpolation (NNI; „Voronoi-Verfahren“) und SIMIK+-Regionalisierung. Geeignete Verfahren müssen insbesondere in der Lage sein, Überschreitungsregionen zu identifizieren, naturräumliche Heterogenität prozessadäquat zu berücksichtigen, und statistische Verteilungseigenschaften abzubilden.

Geostatistische Regressionskriging-Verfahren und darauf aufbauende bedingte geostatistische Simulationen bieten einen etablierten methodischen Werkzeugkasten, der diesen Anforderungen genügt und transparente Entscheidungen ermöglicht. Diese Modelle bestehen aus linearen, potenziell nichtlinear erweiterbaren räumlichen Trendkomponenten sowie Kriging-Komponenten. Sie können z.B. mit Hilfe von Teilmodellen Unterschiede zwischen hydrogeologischen Großräumen berücksichtigen und sind deutschlandweit anwendbar. Der Überschreitungsflächenanteil der Nitratbelastung >50 mg/l kann mit diesen Modellen unverzerrt bestimmt und bei der Entscheidungsfindung einbezogen werden. Sie ermöglichen ferner eine an tatsächlichen Ausweisungsunsicherheiten orientierte Messnetzverdichtung.

Abstract: Evaluation and further development of regionalisation methods for the designation of nitrate-polluted areas according to AVV GeA in Germany

The designation of nitrate-polluted areas for groundwater protection based on the EU Nitrates Directive (91/676/EEC), the Fertiliser Directive and the AVV GeA stipulates the use of geostatistical or deterministic regionalisation methods. The current state-level designation practice is to be replaced by a nationwide geostatistical method. Against this background, this report aims to evaluate the current methods in terms of their suitability and potential for further development, and to investigate the extent to which new or alternative spatial statistical methods are more suitable and could be implemented by the end of 2028.

An in-depth problem analysis and empirical and theoretical model assessment identifies shortcomings and possible biases in the previously used methods of inverse distance weighting (IDW), nearest neighbour interpolation (NNI), and SIMIK+. In particular, suitable methods must be able to identify exceedance regions, take into account spatial heterogeneity, and adequately represent distributional characteristics.

Geostatistical regression-kriging methods and conditional geostatistical simulations offer an established methodological toolbox that fulfils these requirements and enables transparent decisions. These models consist of linear, potentially non-linear spatial trend components and kriging components. With the help of sub-models, for example, these can take into account differences between hydrogeological macro-regions and can be applied throughout Germany. An unbiased estimate of the total exceedance area with nitrate levels >50 mg/l can be obtained from these models and accounted for in the decision-making process. They can also be employed for optimal selection of additional monitoring sites based on actual designation uncertainties.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	10
Tabellenverzeichnis.....	10
Abkürzungsverzeichnis.....	11
Zusammenfassung.....	12
Summary.....	15
1 Einführung.....	18
1.1 Ausgangssituation und Ziele.....	18
1.2 Gliederung und methodisches Vorgehen.....	18
1.3 Unterstützung und Hilfsmittel.....	19
2 Problemanalyse aus geodatenwissenschaftlicher Sicht.....	20
2.1 Grundlagen der Regionalisierung ohne räumlichen Trend.....	20
2.2 Räumliche Trendmodelle.....	22
2.3 Vergleichende Bewertung von Regionalisierungsverfahren.....	24
2.3.1 Qualitative Verfahrensbewertung.....	24
2.3.2 Quantitative Bewertung der Vorhersagegüte.....	24
2.4 Modellierung von Überschreitungsflächen und -wahrscheinlichkeiten.....	27
2.5 Wahl eines Schwellenwertes der Überschreitungswahrscheinlichkeit.....	31
2.6 Spezifische Anforderungen und Herausforderungen.....	32
2.6.1 Nichtstationäre räumliche Abhängigkeit.....	32
2.6.2 Nicht-Normalverteilung.....	34
2.6.3 Bestimmungsgrenze.....	35
2.6.4 Unsicherheiten in erklärenden Variablen.....	35
2.6.5 Skalenwechsel.....	36
2.6.6 Trendbestimmung bei subkritischen Nitratgehalten.....	37
3 Datengrundlagen.....	38
3.1 Messstellen.....	39
3.2 Allgemeine geökologische Raumgliederung.....	39
3.3 Prozessbezogene Umweltvariablen.....	41
3.4 Prozessbezogene Merkmalsextraktion.....	41
4 Bewertung aktuell verwendeter Verfahren und ihrer Weiterentwicklungsmöglichkeiten.....	43
4.1 Voronoi-Verfahren (Nächster-Nachbar-Interpolation, NNI).....	43
4.2 Inverse Distanzgewichtung (IDW).....	44
4.3 G-FlaG.....	45

4.4	SIMIK+	45
5	Empfehlungen für eine bundesweite Regionalisierung	47
5.1	Anforderungen an Regionalisierungsverfahren für die Gebietsausweisung	47
5.2	Empfehlungen hinsichtlich bislang verwendeter Verfahren	47
5.3	Empfehlungen für Verfahrensauswahl und -weiterentwicklung.....	47
5.4	Empfehlungen zu Methoden des maschinellen Lernens.....	48
6	Fazit	50
7	Quellenverzeichnis	51

Abbildungsverzeichnis

Abbildung 1:	Regionalisierung von Nitratkonzentrationen im Pilotgebiet mit Ordinary Kriging (OK) und Kriging mit externer Drift (KED)21
Abbildung 2:	Deutschlandweite Semivariogramme der Nitratkonzentration sowie der Residuen eines linearen Trendmodells.....22
Abbildung 3:	Histogramme von Vorhersageentfernungen in Leave-one-out-Kreuzvalidierung, Zehnfach-Kreuzvalidierung und bei bundesweiter Regionalisierung25
Abbildung 4:	Vorhersagefehlerprofile für NNI, IDW, OK, KED, MLR und RF..26
Abbildung 5:	Vorhersagefehlerprofile für die Über-/Unterausweisung nitratbelasteter Flächen mit NNI, IDW, OK, KED, MLR und RF.28
Abbildung 6:	Ergebnisse bedingter geostatistischer Simulationen basierend auf Messungen an drei fiktiven Messstellen30
Abbildung 7:	Regionalisierung der Überschreitungswahrscheinlichkeit für 50 mg/l im Pilotgebiet mit bedingter geostatistischer Simulation30
Abbildung 8:	Bestimmung des Überschreitungsflächenanteils im Pilotgebiet und seine statistische Unsicherheit basierend auf bedingter geostatistischer Simulation31
Abbildung 9:	Abnahme des ermittelten Überschreitungsflächenanteils im Pilotgebiet mit zunehmendem Schwellenwert ϕ der Überschreitungswahrscheinlichkeit32
Abbildung 10:	Semivariogramme der Regressionsresiduen innerhalb über GWK-Grenzen hinweg sowie innerhalb desselben GWK33
Abbildung 11:	Semivariogramme der Regressionsresiduen innerhalb ausgewählter hydrogeologischer Großräume.....34
Abbildung 12:	Regionalisierung von Nitratkonzentrationen in mg/l im Pilotgebiet mit Nächster-Nachbar-Interpolation (NNI, Voronoi-Verfahren) und Inverser Distanzgewichtung (IDW)44
Abbildung 13:	Einfluss der Lage von Messstellen bei IDW und OK45

Tabellenverzeichnis

Tabelle 1:	Bestimmung des Überschreitungsflächenanteils im Pilotgebiet mit NNI, IDW, OK, KED, RF und BGS29
Tabelle 2:	Gegenüberstellung von Kartenmaßstab, kleinster kartierbarer Einheit und empfohlener Raster-Auflösung für Geodaten36
Tabelle 3:	Datengrundlagen: 1. Geoökologische Raumgliederung.....38
Tabelle 4:	Datengrundlagen: 2. Prozessbezogene Umweltvariablen40
Tabelle 5:	Datengrundlagen: 3. Prozessbezogene Merkmalsextraktion ..41

Abkürzungsverzeichnis

AEA	<i>Agri-environmental archetypes</i>
BGR	Bundesanstalt für Geowissenschaften und Rohstoffe
BGS	Bedingte geostatistische Simulation
BKG	Bundesamt für Kartographie und Geodäsie
BKR	Boden-Klima-Räume
DEM	<i>Digital elevation model</i>
DGM	Digitales Geländemodell
EZG	Einzugsgebiet
FABDEM	<i>Forest and Buildings removed Copernicus DEM</i>
GAM	<i>Generalised additive model</i>
G-FlaG	Geostatistische Flächenanteilsschätzung zur Zustandsbewertung von GWK
GWK	Grundwasserkörper
GWN	Grundwasserneubildungsrate
HÜK	Hydrogeologische Übersichtskarte
HYRAUM	Hydrogeologische Raumgliederung von Deutschland
IDW	Inverse Distanzgewichtung
KED	Kriging mit externer Drift (<i>Kriging with external drift</i>)
KV	Kreuzvalidierung
LBM	Digitales Landbedeckungsmodell
MLR	Multiple lineare Regression
OK	<i>Ordinary Kriging</i>
RF	<i>Random Forest</i>
RFRK	<i>Random Forest Regression Kriging</i>
RMSE	<i>Root mean-square error</i>

Zusammenfassung

Als Grundlage für einen verbesserten Grundwasserschutz erfordert die EU-Nitratrichtlinie (91/676/EWG) in Verbindung mit der Düngeverordnung die Ausweisung nitratbelasteter Gebiete. Die Gebietsausweisung erfolgt gemäß der Allgemeinen Verwaltungsvorschrift zur Ausweisung von mit Nitrat belasteten und eutrophierten Gebieten (AVV GeA) anhand von Messwerten an derzeit rund 13.500 Grundwassermessstellen. Hierfür sind in Abhängigkeit von der Messstellendichte deterministische Verfahren (Inverse Distanzgewichtung, IDW, sowie Voronoi-Verfahren bzw. Nächster-Nachbar-Interpolation, NNI) oder geostatistische Regionalisierungsverfahren zu verwenden. Die aktuelle Ausweisungspraxis auf Landesebene soll in ein bundesweit einheitliches geostatistisches Regionalisierungsverfahren überführt werden.

Das vorliegende Gutachten verfolgt vor diesem Hintergrund die folgenden Ziele:

1. Die bisherigen Verfahren zur Gebietsausweisung sollen hinsichtlich ihrer Eignung zur flächenhaften Abgrenzung von Nitratreintragsgebieten aus fachlicher Sicht evaluiert und bewertet werden.
2. Es soll eruiert werden, ob und wenn ja, wie die bisherigen Verfahren weiterentwickelt werden könnten, um die Flächenabgrenzung unter den gegebenen Randbedingungen maximal robust und verursachergerecht vorzunehmen.
3. Über die bisherigen Verfahren zur Gebietsausweisung hinaus soll untersucht werden, inwieweit neue oder alternative Methoden der räumlichen Statistik besser geeignet sind. Dabei soll auch eine Empfehlung abgegeben werden, wie mit der verpflichtenden Einführung geostatistischer Verfahren bis Ende 2028 umzugehen ist und wie diese umgesetzt werden könnte.

Die Problemanalyse in Abschnitt 2 formalisiert die Aufgabe der Gebietsausweisung als Überschreitungsregionsproblem, welches im Gegensatz zur herkömmlichen Betrachtung als Regionalisierungsproblem steht, und zeigt damit eine mögliche Quelle erheblicher Verzerrungen auf. Analysen in einer Pilotregion sowie deutschlandweit dienen der Veranschaulichung. Grundsätzlich bieten Regionalisierungsmodelle und insbesondere das Regressionskriging (Kriging mit externer Drift, KED) umfassende Gestaltungsmöglichkeiten für die Modellierung von räumlicher Heterogenität mit Hilfe flächenhaft verfügbarer Zusatzinformationen. Aus der Problemanalyse werden anschließend die quantitativen Bewertungskriterien und qualitativen Anforderungen an datengetriebene Modelle der Nitratbelastung abgeleitet. Die deterministischen Verfahren der AVV GeA (IDW, NNI) erfüllen diese Kriterien im Vergleich zu KED nicht, und das SIMIK+-Regressionskriging weist vermeidbare Einschränkungen und Mängel auf. Die bedingte geostatistische Simulation (BGS) wird als Verfahren eingeführt, welches Wahrscheinlichkeiten der Überschreitung des Nitratschwellenwertes regionalisieren kann und es ferner erlaubt, die tatsächlich nitratbelastete Gesamtfläche unverzerrt zu berechnen. Die Charakterisierung der Gebietsausweisung als Diagnoseproblem eröffnet hierbei den Blick auf eine Abwägung der Folgen falsch-positiver gegenüber falsch-negativen Gebietsausweisungen. Des Weiteren werden technische Aspekte wie die Modellierung nichtstationärer Trends und Abhängigkeiten sowie nicht-normalverteilter Daten adressiert und die Potenziale von Modellen des maschinellen Lernens diskutiert. Die Kombination mit dem sekundären Ausweisungskriterium, dem Vorliegen eines ansteigenden Trends bei Nitratgehalten über 37,5 mg/l, wird abschließend skizziert.

Flächenhaft vorliegende Datengrundlagen werden anschließend hinsichtlich ihrer Eignung als Zusatzinformationen oder erklärende Variablen in der Regionalisierung der Nitratbelastung untersucht (Abschnitt 3). Dabei können allgemeine geoökologische Raumgliederungen von prozessbezogenen Umweltvariablen unterschieden werden. Eine dritte Gruppe sind abgeleitete prozessbezogene Merkmale, wie beispielsweise die Einzugsgebietsmittelwerte erklärender

Variablen (z.B. landwirtschaftlicher Flächenanteil im Anstrombereich). Eine erkundende Analyse der Daten weist insbesondere auf einen Nutzen von Boden-Klima-Räumen, hydrogeologischen Großräumen, geochemischen Gesteinstypen und landwirtschaftlichen Flächenanteilen im Einzugsgebiet hin. Eine separate Berücksichtigung dieser Variablen innerhalb der hydrogeologischen Großräume (Interaktionsterme) erhöht die erklärte Varianz substantiell. Die Zusatzinformationen liegen jedoch teils in groben Detaillierungsgraden vor (Maßstäbe bis 1:1.000.000 und Auflösungen bis 1 km × 1 km bzw. Gemeindeebene), woraus sich Unsicherheiten vor allem entlang von Zonengrenzen ergeben. Grundsätzlich ist vor diesem Hintergrund der in der AVV GeA geforderte Zielmaßstab von 1:25.000 kritisch zu hinterfragen.

Aufbauend auf den vorangegangenen Abschnitten werden in Abschnitt 5 Anforderungen an Regionalisierungsverfahren formuliert:

1. Regionalisierungsverfahren müssen in der Lage sein, Überschreitungs-Wahrscheinlichkeiten – und nicht lediglich erwartete Nitratgehalte – räumlich vorherzusagen. Adäquate Verteilungsmodelle sind hierfür erforderlich.
2. Sie müssen hierbei in der Lage sein, erklärende Variablen unterschiedlicher Merkmalstypen einzubeziehen, um räumliche Muster und Trends prozessadäquat und verursachergerecht zu berücksichtigen.
3. Da heterogene (nichtstationäre) räumliche Abhängigkeiten nachgewiesen wurden, sind diese adäquat zu modellieren.
4. Des Weiteren ist gegebenenfalls der Skalenwechsel von Punktmessungen zu Kleinflächen (Rasterzellen) zu modellieren, wobei Zielmaßstab oder -auflösung kritisch hinterfragt werden sollten.

Es wird darauf hingewiesen, dass es für einen ausgewogenen Grundwasserschutz weiterhin der Bestimmung eines geeigneten Überschreitungswahrscheinlichkeits-Entscheidungswertes bedarf. Die AVV GeA trifft hierzu keine spezifischen Festlegungen.

Auf Grundlage dieser Anforderungen ergeben sich folgende Schlussfolgerungen hinsichtlich der Ziele des Gutachtens. Zu Ziel 1 wird festgestellt, dass von den in der AVV GeA betrachteten Verfahren lediglich die geostatistischen Methoden den Anforderungen genügen. Eine Verwendung von IDW oder NNI (Voronoi-Verfahren) wird nicht empfohlen, und zwar unabhängig von den in der AVV GeA formulierten, für nicht zweckmäßig erachteten Kriterien für die Messstellendichte. Das einzige bisher verwendete geostatistische Verfahren, SIMIK+, weist ferner Einschränkungen und technische Fehler auf; allgemeinere, gut etablierte geostatistische Verfahren (KED, BGS) existieren, weshalb eine Weiterentwicklung nicht sinnvoll erscheint (Ziel 2).

Im Hinblick auf Ziel 3 des Gutachtens wird eine Vorgehensweise aus dem Bereich der geostatistischen Regionalisierung vorgeschlagen, die etablierte Werkzeuge nutzt und auch kurzfristig deutschlandweit umsetzbar ist. Dieser nutzt im Kern BGS oder KED für die Abschätzung von Überschreitungswahrscheinlichkeiten und -flächenanteil. Ein lineares, ggf. nichtlineares räumliches Trendmodell erfasst dabei prozessbezogene Muster und räumliche Abhängigkeiten, wobei Teilmodelle innerhalb hydrogeologischer Großräume regionale Heterogenität berücksichtigen können. Von einer separaten Modellierung der Nitratbelastung auf Bundeslandebene wird dagegen abgeraten. Datentransformationen, Merkmalsextraktion und datengetriebene Modellselektion sind weitere erforderliche Schritte der Modellbildung. Die Frage eines möglichen Skalenwechsels von der Punkt- zur Rasterzellen-Ebene in geeigneter Auflösung ist zu klären und kann durch Block-Geostatistik umgesetzt werden.

Es wird weiterhin empfohlen, die genutzten geostatistischen Modelle auch zur optimalen Standortauswahl in der weiteren Verdichtung des Messstellennetzes zu nutzen, um mit dem erforderlichen Ressourceneinsatz den bestmöglichen Nutzen zu erzielen.

Hinsichtlich des Potenzials von Verfahren des maschinellen Lernens, deren Nutzung jedoch derzeit nicht explizit durch die AVV GeA abgesichert ist, muss konstatiert werden, dass derzeit keine belastbaren Hinweise auf verbesserte räumliche Vorhersagen vorliegen und die eingeschränkten Interpretationsmöglichkeiten von Black-Box-Modellen neue Fragen hinsichtlich der Nachvollziehbarkeit von Entscheidungen aufwerfen. Weitere Untersuchungen und Methodenentwicklungen im Hinblick auf längerfristige Nutzungsperspektiven werden empfohlen.

Abschließend ist zu konstatieren, dass die die Ausweisung nitratbelasteter Gebiete spezielle Anforderungen an Regionalisierungsverfahren stellt, und ihr unreflektierter Einsatz in systematischen Unter- oder Überausweisungen resultieren kann. Das weite Feld der geostatistischen Regressionskriging-Verfahren bietet einen etablierten methodischen Werkzeugkasten für eine prozessadäquate Modellierung und transparente Entscheidungsfindung.

Summary

As a basis for improved groundwater protection, the EU Nitrates Directive (91/676/EEC) in conjunction with the German Fertiliser Directive (Düngeverordnung) requires the designation of nitrate-polluted areas. Areas are designated in accordance with the General Administrative Regulation on the Designation of Nitrate-Polluted and Eutrophic Areas (AVV GeA) on the basis of measurements at currently around 13,500 groundwater monitoring wells. Depending on the density of measurements, deterministic methods (inverse distance weighting, IDW, and nearest neighbour interpolation, NNI) or geostatistical regionalisation methods are to be used for this purpose. The current state-level designation practice is to be replaced by a nationwide geostatistical regionalisation.

Against this background, this expert report pursues the following objectives:

1. to evaluate and assess the existing procedures for area designation with regards to their suitability for the delineation of nitrate input areas from a technical perspective.
2. to determine whether and, if so, how the existing procedures could be further developed in order to maximise the robustness and adequacy of the area designation under the given boundary conditions.
3. beyond the existing methods for area designation, it should be investigated to what extent new or alternative spatial statistical methods are more suitable. A recommendation will also be made on how to deal with the mandatory introduction of geostatistical methods by the end of 2028 and how this could be implemented.

The problem analysis in section 2 formalises the task of area designation as an exceedance region problem, which is in contrast to the conventional treatment as a regionalisation problem, and thus identifies a possible source of considerable bias. Analyses in a pilot region and Germany-wide analyses serve to illustrate this. In principle, regionalisation models and in particular regression kriging (kriging with external drift, KED) offer extensive design options for modelling spatial heterogeneity with the help of ancillary information available areawide. The quantitative evaluation criteria and qualitative requirements for data-driven models of nitrate pollution are then derived from the problem analysis. The deterministic methods of the AVV GeA (IDW, NNI) do not fulfil these criteria in comparison to KED, and SIMIK+ regression kriging has avoidable limitations and shortcomings. Conditional geostatistical simulation (CGS) is introduced as a method that can regionalise probabilities of exceeding the nitrate threshold and also allows the unbiased estimation of the total nitrate-polluted area. The characterisation of area designation as a diagnostic problem opens up the possibility of finding a trade-off between false-positive and false-negative designations. Furthermore, technical aspects such as the modelling of non-stationary trends and dependencies as well as non-normally distributed data are addressed and the potential of machine-learning models is discussed. Finally, the combination with the secondary designation criterion, referring to increasing trends in areas with nitrate levels above 37.5 mg/l, is outlined.

Data bases available for specific areas are then examined with regard to their suitability as ancillary information or explanatory variables in the regionalisation of nitrate pollution (section 3). General geo-ecological spatial classifications can be distinguished from process-related environmental variables. A third group consists of derived process-related characteristics, such as the catchment mean values of explanatory variables (e.g. agricultural area in the upstream area). An exploratory analysis of the data indicates in particular the utility of soil-climate regions, hydrogeological macro-regions, geochemical rock types and the fraction of agricultural land in the catchment area. Using these variables within the hydrogeological regions (as interaction terms) substantially increases the explained variance. However, some of the ancillary information is

only available at a coarse level of detail (scales of up to 1:1,000,000 and resolutions of up to 1 km × 1 km or municipality level), which results in uncertainties, particularly along zone boundaries. Against this background, the target scale of 1:25,000 required in the AVV GeA must be critically scrutinised.

Building on the previous sections, general requirements for regionalisation procedures are formulated in section 5:

1. Regionalisation procedures must be able to spatially predict exceedance probabilities – and not merely expected nitrate concentrations. For this, distributional assumptions must be satisfied.
2. They must be able to include explanatory variables of different data types in order to account for spatial patterns and trends in a process-adequate and cause-appropriate manner.
3. Since heterogeneous (non-stationary) spatial dependencies have been identified, these must be adequately modelled.
4. Furthermore, the change of spatial support from point measurements to small areal units (raster cells) may need to be modelled, whereby the target scale or resolution should be critically scrutinised.

It should be noted that the determination of a suitable exceedance-probability decision threshold is required for balanced groundwater protection. The AVV GeA does not make any specific stipulations in this regard.

Based on these requirements, the following conclusions can be drawn with regard to the report's objectives. Concerning Objective 1, it is established that of the methods considered in the AVV GeA, only the geostatistical methods fulfil the requirements. The use of IDW and NNI is not recommended, irrespective of the criteria formulated in the AVV GeA for the required density of the monitoring network, which are deemed unsuitable. The only geostatistical method used to date, SIMIK+, also has limitations and technical shortcomings; more general, well-established geostatistical methods (KED, CGS) exist, which is why further development of the other does not appear to be warranted (Objective 2).

Regarding Objective 3 of the report, a geostatistical regionalisation approach is proposed that uses established tools and can also be applied nationwide in the short term. This approach can be built around CGS or KED for the estimation of exceedance probabilities and total exceedance area. A linear or possibly non-linear spatial trend model is used to capture process-related patterns and spatial dependencies, whereby sub-models can be built within hydrogeological macro-regions in order to take regional heterogeneity into account. Separate modelling of nitrate pollution at the state level, on the other hand, is not recommended. Data transformations, feature extraction and data-driven model selection are additional necessary steps in model construction. The question of a possible change of support from point scale to raster cells at a suitable resolution must be clarified and can be implemented using block geostatistics.

It is further recommended that the geostatistical models used should also be employed for optimal selection of additional monitoring sites in order to achieve the best possible benefit with the available resources.

Regarding the potential of machine learning methods, the use of which is currently not explicitly supported by the AVV GeA, it must be noted that there is currently no reliable evidence of an improved predictive performance, and the limited interpretability of black-box models moreover raises new questions concerning the traceability of decisions. Further research is recommended in order to assess longer-term potentials of these methods.

In conclusion, it can be stated that the designation of nitrate-polluted areas places special demands on regionalisation procedures, and their inappropriate use can result in a systematic under- or over-designation of these areas. The broad field of geostatistical regression-kriging methods offers an established methodological toolbox for process-adequate modelling and transparent decision-making.

1 Einführung

1.1 Ausgangssituation und Ziele

Erhöhte Nitratgehalte im Grundwasser sind ein globales Umwelt- und Gesundheitsproblem, welches vor allem in landwirtschaftlich intensiv genutzten Regionen auftritt (Erisman et al., 2013). Vor diesem Hintergrund enthalten die EU-Nitratrichtlinie (91/676/EWG) sowie daraus abgeleitete nationale Bestimmungen der Düngeverordnung Regelungen zur Ausweisung nitratbelasteter und eutrophierter Gebiete und Verminderung der Nitratreinträge insbesondere aus landwirtschaftlicher Bewirtschaftung. Konkrete Vorgaben zur technischen Umsetzung einer Ausweisung solcher Gebiete wurden in Form der Allgemeinen Verwaltungsvorschrift zur Ausweisung von mit Nitrat belasteten und eutrophierten Gebieten (AVV GeA) erlassen.

Diese Gebietsausweisung erfolgt anhand von Messwerten an derzeit rund 13.500 Grundwassermessstellen, wobei die AVV GeA in Abhängigkeit von der Messstellendichte deterministische (Inverse Distanzgewichtung – IDW, Voronoi-Verfahren bzw. Nächster-Nachbar-Interpolation, NNI) sowie geostatistische Regionalisierungsverfahren zulässt bzw. vorschreibt. Maßgebliche Kriterien sind die Überschreitung eines Schwellenwertes der Nitratkonzentration von 50 mg/l im Vierjahresmittel (der Jahreshöchstwerte) oder das Vorliegen eines steigenden (Fünfjahres-) Trends bei Überschreitung von 37,5 mg/l, mit Ausnahme bestimmter Sondersituationen. Derzeit führen die Bundesländer die Gebietsausweisungen unabhängig voneinander und mit unterschiedlichen Verfahren durch. Eine Vorgabe der EU-Kommission ist jedoch, bundesweit ein einheitliches Verfahren einzuführen und hierfür die Anforderungen an eine Messstellendichte zu erfüllen, die gemäß AVV GeA für die Anwendung eines geostatistischen Regionalisierungsverfahrens erforderlich ist (UBA, 2023).

Vor diesem Hintergrund verfolgt das vorliegende, durch das Umweltbundesamt in Auftrag gegebene Sachverständigengutachten die folgenden Ziele (UBA, 2023):

1. Die bisherigen Verfahren zur Gebietsausweisung sollen hinsichtlich ihrer Eignung zur flächenhaften Abgrenzung von Nitratreintragsgebieten aus fachlicher Sicht evaluiert und bewertet werden.
2. Es soll eruiert werden, ob und wenn ja, wie die bisherigen Verfahren weiterentwickelt werden könnten, um die Flächenabgrenzung unter den gegebenen Randbedingungen maximal robust und verursachergerecht vorzunehmen.
3. Über die bisherigen Verfahren zur Gebietsausweisung hinaus soll untersucht werden, inwieweit neue oder alternative Methoden der räumlichen Statistik besser geeignet sind, eine problemadäquate Abgrenzung von Flächen zu erreichen.
Dieser Teil des Gutachtens sollte auch eine Empfehlung dazu enthalten, wie mit der verpflichtenden Einführung geostatistischer Verfahren bis Ende 2028 umzugehen ist und Vorschläge zur konkreten Umsetzung enthalten.

1.2 Gliederung und methodisches Vorgehen

Zur Beantwortung der vorgegebenen Fragestellungen beginnt dieses Gutachten mit einer konzeptionellen, mathematischen und empirisch-statistischen Problemanalyse mit Blick auf die Identifikation von Überschreitungsflächen mit Hilfe von Regionalisierungsverfahren (Abschnitt 2). Hieraus leiten sich die Anforderungen an datengetriebene Modelle der Nitratbelastung sowie Kriterien für Modellvergleiche ab, welche auf bisher verwendete wie auch andere in Frage kommende Verfahren angewendet werden. Über die zu betrachtenden deterministischen und geostatistischen Regionalisierungsverfahren hinaus werden auch neuere Verfahren des maschinellen Lernens sowie kombinierte Verfahren diskutiert. Möglichen Vorgehen zur Berücksichtigung

des sekundären Ausweiskriteriums – Vorliegen eines ansteigenden Trends bei Nitratgehalten über 37,5 mg/l – wird ein Unterabschnitt gewidmet (Abschnitt 2.6.6).

Die Identifikation und Bewertung geeigneter Datengrundlagen für die Regionalisierung wird als zentraler Bestandteil der Modellbildung angesehen und daher in einem eigenen Abschnitt betrachtet (Abschnitt 3).

Mit dieser Hintergrundinformation werden anschließend in Abschnitt 4 die bisher verwendeten Verfahren zur Gebietsausweisung bewertet (Ziel 1), hinsichtlich ihrer Weiterentwicklungsmöglichkeiten beurteilt (Ziel 2) und in Abschnitt 5 Empfehlungen für neue oder alternative Methoden der räumlichen Statistik für eine bundesweite Gebietsausweisung vorgestellt (Ziel 3).

Es sei angemerkt, dass es über die Methodenbewertung und -weiterentwicklung hinaus nicht das Ziel dieser Arbeit ist, einen finalen, technischen Workflow für eine Gebietsausweisung vorzuschlagen oder Gebiete für eine Ausweisung als belastete Gebiete zu präjudizieren. In diesem Sinne dienen die im Gutachten enthaltenen kartographischen Darstellungen lediglich der Veranschaulichung der Methoden und ihrer Vor- und Nachteile; sie sind nicht dazu geeignet, einzelne Flächen als nitratbelastet auszuweisen oder ihnen die empirische Grundlage für eine Nitratbelastung abzusprechen. Kartographische Darstellungen enthalten aus diesem Grund absichtlich keine konkrete geographische Referenz. In diesem Zusammenhang sei auch darauf hingewiesen, dass im Rahmen dieser Arbeit eine umfassende Datenvorverarbeitung und Qualitätsprüfung der Messdaten nicht möglich war (u.a. Ausreißerentfernung, hydrogeologische Plausibilitätsprüfung), da dies u.U. eine zeitaufwendige Einzelfallbeurteilung unter Heranziehung lokalen hydrogeologischen Expertenwissens erfordert.

1.3 Unterstützung und Hilfsmittel

Die Erstellung dieses Sachverständigengutachtens wäre ohne die Unterstützung durch zahlreiche Personen nicht möglich gewesen – ihnen allen gilt mein herzlicher Dank:

- ▶ Herr Falk Hilliges und Herr Matthias Rothe als die fachlichen Begleiter, die wertvolle Hinweise gaben und zusammen mit Herrn Niels Richter erforderliche Daten zur Verfügung stellten;
- ▶ Herr Dr. Thomas Süße und Herr Dr. Manfred Fink, die diese Begutachtung mit ihrem Fachwissen in räumlicher Statistik bzw. Ökohydrologie begleiteten, für ihre wertvollen Hinweise und Beiträge;
- ▶ den Fachleuten der Länder, die im Rahmen des Workshops am 27.10.2023 relevante Hinweise gaben, sowie Frau Annett Peters (Thüringer Landesamt für Umwelt, Bergbau und Naturschutz) für den offenen und transparenten Austausch;
- ▶ den Datenerzeugern, hier insbesondere dem AGRUM-DE-Team, für die Weitergabe von Daten für diesen Zweck.

Weiterhin wurde R (Version 4.2.1) mit den Erweiterungen (*packages*) *car*, *gam*, *gstat*, *randomForest*, *RSAGA*, *sf*, *sperrorest*, *terra* und weiteren Modulen verwendet. Englischsprachige Zusammenfassungen basieren auf DeepL.com-Übersetzungen.

2 Problemanalyse aus geodatenwissenschaftlicher Sicht

Die folgenden Abschnitte stellen den konzeptionellen und methodischen Rahmen der Ausweisung von nitratbelasteten Flächen vor. Hierzu werden zunächst grundlegende Konzepte der Regionalisierung quantitativer regionalisierter Variablen eingeführt und formalisiert (Abschnitt 2.1). Diese werden anschließend um lineare und nichtlineare Trendmodelle erweitert, die es erlauben, Zusatzkriterien wie Hydrogeologie und Landbedeckung in Regionalisierungsmodellen zu berücksichtigen (Abschnitt 2.2). Im nächsten Abschnitt werden Bewertungskriterien für Regionalisierungsverfahren erörtert (Abschnitt 2.3).

Unter Beachtung dieser Kriterien wird das Regionalisierungsproblem anschließend als Modellierung von Überschreitungswahrscheinlichkeiten und -flächen reinterpretiert (Abschnitt 0). Hieraus ergibt sich die Notwendigkeit der Bestimmung eines Schwellenwertes der Überschreitungswahrscheinlichkeit (Abschnitt 2.5).

Einige wichtige, aber eher technische Aspekte sind in Abschnitt 2.6 ausgelagert, um in den vorgenannten Abschnitten zunächst die übergeordneten Aspekte zu diskutieren. Die Leserinnen und Leser mögen Verständnis dafür haben, dass hierdurch Konzepte wie die Box-Cox-Transformation, auf die bereits vorher Bezug genommen wird, erst nachgelagert erläutert werden.

2.1 Grundlagen der Regionalisierung ohne räumlichen Trend

Die Regionalisierung von Punktmessungen einer quantitativen Variable in einer Region D wird wahrscheinlichkeitstheoretisch durch ein räumliches Zufallsfeld $Z = (Z(\mathbf{x}))_{\mathbf{x} \in D}$ repräsentiert, dessen Realisierungen, d.h. konkreten Werte

$$z_1 = z(\mathbf{x}_1), \dots, z_n = z(\mathbf{x}_n),$$

an n Messtellen $\mathbf{x}_1, \dots, \mathbf{x}_n \in D$ beobachtet werden. Diese Werte werden als regionalisierte Werte bezeichnet und bilden die Lernstichprobe des Regionalisierungsverfahrens.

Ziel der Regionalisierung ist es, an einer beliebigen Stelle $\mathbf{x}_0 \in D$, an der die Realisierung $z_0 = z(\mathbf{x}_0)$ unbekannt ist, durch einen Prädiktor $\hat{Z}(\mathbf{x}_0)$ vorherzusagen. Diese Stelle \mathbf{x}_0 ist dabei nur stellvertretend z.B. für alle Vorhersagestellen auf einem Gitternetz auf D . Wird der Prädiktor als gewichtete Summe der gemessenen Zufallsvariablen gebildet, wird er als linear bezeichnet:

$$\hat{Z}(\mathbf{x}_0) = \lambda_1 Z(\mathbf{x}_1) + \dots + \lambda_n Z(\mathbf{x}_n).$$

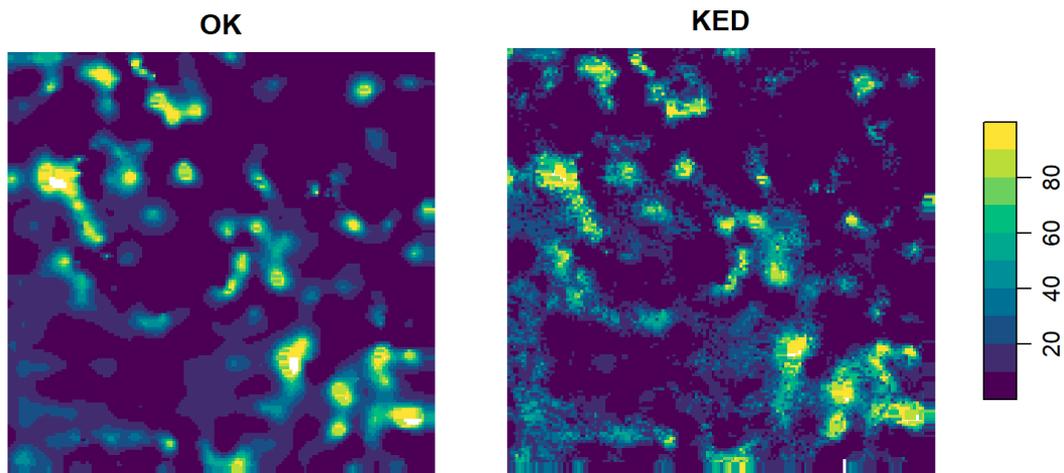
Neben dem geostatistischen Verfahren des Ordinary Kriging (OK; Abbildung 1) lassen sich auch IDW und NNI (vgl. Abschnitt 4) auf diese Weise darstellen, grundsätzlich jedoch auch andere Verfahren, die nur die regionalisierten Werte für eine Interpolation heranziehen.

Ein Verfahren, das im Mittel (über Realisierungen) mit dem wahren Erwartungswert $m(\mathbf{x}_0) = E(Z(\mathbf{x}_0))$ übereinstimmt, d.h.

$$E(\hat{Z}(\mathbf{x}_0)) = m(\mathbf{x}_0)$$

wird als unverzerrt (*unbiased*) bezeichnet. Diese (grundsätzlich erstrebenswerte) Eigenschaft ist für OK, IDW und NNI erfüllt, sofern das Zufallsfeld Z nicht systematisch in Richtung höherer oder niedrigerer Werte „driftet“, d.h. sofern $E(Z(\mathbf{x}) - Z(\mathbf{y})) = 0$ für alle Punkte $\mathbf{x}, \mathbf{y} \in D$.

Abbildung 1: Regionalisierung von Nitratkonzentrationen im Pilotgebiet mit Ordinary Kriging (OK) und Kriging mit externer Drift (KED)



Die Regionalisierungen wurden für Box-Cox-transformierte Nitratkonzentrationen durchgeführt und anschließend zu Konzentrationen in mg/l rücktransformiert.

Quelle: eigene Darstellung, A. Brenning.

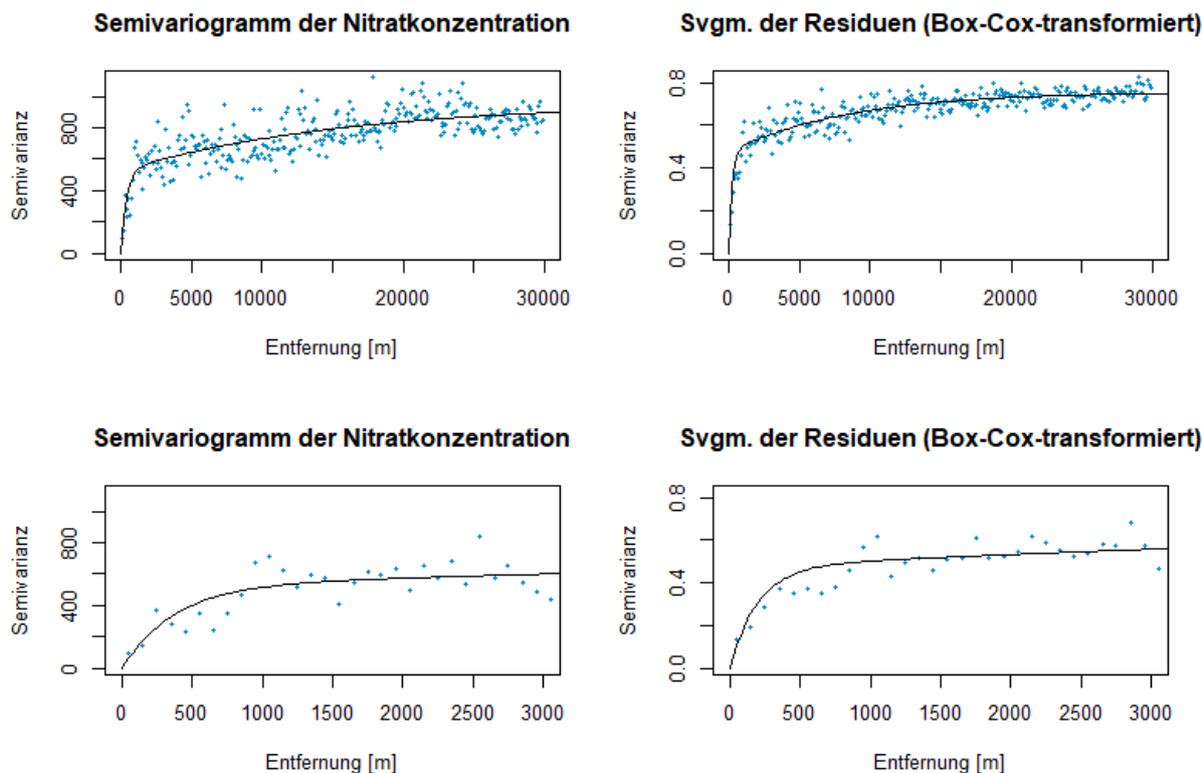
Der Fehler im quadratischen Mittel des Prädiktors, oder die Vorhersagevarianz, wird mit

$$\sigma^2(\mathbf{x}_0) = E \left(\left(\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0) \right)^2 \right)$$

bezeichnet. Sofern die Voraussetzungen von OK erfüllt sind, ist OK derjenige unverzerrte Prädiktor, der diesen Fehler minimiert. OK wird daher als bester unverzerrter linearer Prädiktor bezeichnet. Sind die Voraussetzungen erfüllt, können z.B. IDW oder NNI somit nicht optimal sein, da sie offensichtlich nicht mit OK übereinstimmen.

Eine Voraussetzung für die Optimalität geostatistischer Prädiktoren ist dabei, dass die räumliche Abhängigkeit korrekt modelliert wird, wozu ein Semivariogramm verwendet wird und (eher aus praktischen als aus mathematischen Gründen) in der Regel Stationaritätsannahmen getroffen werden. Diese besagen (vereinfacht gesagt), dass die Stärke der räumlichen Abhängigkeit überall in der Region D gleich ist, und gleich schnell mit zunehmender Entfernung zwischen zwei Messstellen abnimmt. Diese Stationarität des Semivariogramms ist somit eine Homogenitätsannahme, die grundsätzlich überprüfbar und „aufweichbar“ ist (Brenning, 2001; siehe auch Abschnitt 2.6.1). Die in Abbildung 2 dargestellten deutschlandweiten Semivariogramme der Nitratkonzentration stellen eine solche stationäre Abhängigkeit vereinfachend dar. Hier ist zu erkennen, dass bereits über kurze Distanzen (<1 km) etwa 40% der Gesamtvarianz von rund 1200 $(\text{mg/l})^2$ auftreten. Dies ist zwar ein Zeichen hoher lokaler Heterogenität, jedoch besteht selbst über Entfernungen >3 km noch eine räumliche Abhängigkeit in Höhe von etwa 50% der Gesamtvarianz. Dies ist kein schlechter Ausgangspunkt für eine Regionalisierung, weist jedoch auch auf bedeutende Unsicherheiten in größeren Lücken im Messstellennetz hin.

Abbildung 2: Deutschlandweite Semivariogramme der Nitratkonzentration sowie der Residuen eines linearen Trendmodells



Betrachtet werden die Semivariogramme der Residuen von Trendmodellen für die Box-Cox-transformierten Daten. An die empirischen Semivariogramme wurden jeweils zwei aufaddierte exponentielle Semivariogramm-Modelle angepasst, die unterschiedliche räumliche Skalen der Abhängigkeit repräsentieren. Die deutschlandweite Gesamt-Varianz der Nitratkonzentrationsmessungen beträgt 1201 (mg/l)^2 . Die Einheit der Semivarianzen der (untransformierten) Nitratkonzentration ist ebenfalls (mg/l)^2 .

Quelle: eigene Darstellung, A. Brenning.

2.2 Räumliche Trendmodelle

Die OK-Annahme einer fehlenden Drift, d.h. $E(Z(\mathbf{x}) - Z(\mathbf{y})) = 0$ für alle Punkte $\mathbf{x}, \mathbf{y} \in D$, ist vielfach, und so auch bei der räumlichen Verteilung von Nitratkonzentrationen im Grundwasser, verletzt, da $m(\mathbf{x})$ von externen Faktoren abhängt („externe Drift“). Diese können in einem räumlichen Trendmodell als erklärende Variablen $f_1(\mathbf{x}), \dots, f_p(\mathbf{x})$ genutzt werden, um die Regionalisierung zu verbessern, sofern ihre Werte flächenhaft im gesamten Gebiet D bekannt sind. Sind diese Beziehungen linear,

$$Z(\mathbf{x}) = \alpha + \beta_1 f_1(\mathbf{x}) + \dots + \beta_p f_p(\mathbf{x}) + e(\mathbf{x}),$$

so ist Universal Kriging oder Kriging mit externer Drift (KED) der beste unverzerrte lineare Prädiktor von $Z(\mathbf{x}_0)$ (siehe Abbildung 1). Die räumliche Abhängigkeit des Residuen-Zufallsfeldes $(e(\mathbf{x}))_{\mathbf{x} \in D}$ wird dabei durch ein (in der Regel stationäres) Semivariogramm modelliert.

KED zieht somit seine Vorhersagekraft zum einen aus dem „globalen“ linearen Trendmodell, dessen Güte beispielsweise durch die erklärte Varianz ausgedrückt werden kann; zum anderen nutzt es die räumliche Autokorrelation der Residuen, um die unerklärte Komponente zu regionalisieren.

Bei der linearen räumlichen Trendmodellierung mit der in Abschnitt 3 beschriebenen Datengrundlage werden deutschlandweit 21% der Varianz der (Box-Cox-transformierten; siehe Abschnitt 2.6.2) Nitratkonzentration erklärt ($R^2_{adj} = 0,21$), wobei innerhalb der einzelnen hydrogeologischen Großräume zwischen 10 und 59% der Varianz erklärt werden (am niedrigsten im Mitteldeutschen Bruchschollenland sowie im Nord-/mitteldeutschen Lockergestein). Auf Grundlage des Semivariogramms der Residuen in Abbildung 2 kann grob abgeschätzt werden, dass auf ca. 1-3 km Entfernung von Messstellen in etwa ein weiteres Drittel bis ein Viertel der Gesamtvarianz durch die Interpolation der Modellfehler erklärt werden kann, zusätzlich zu dem durch das lineare Modell erklärten Varianzanteil.

Die lineare Form des Trendmodells erscheint zunächst als starke Einschränkung, jedoch wird bei Betrachtung des umfangreichen Werkzeugkastens der linearen Modellierung schnell deutlich, dass beispielsweise auch nichtlineare Transformationen der erklärenden Variablen denkbar sind oder (lineare) Interaktionsterme (oder Wechselwirkungen) hinzugefügt werden können, um kontrastierende Beziehungen in unterschiedlichen Teilgebieten zu berücksichtigen. Beispielsweise sind bei der Regionalisierung der Nitratgehalte Interaktionen mit hydrogeologischen Großräumen zumindest teilweise sinnvoll, da diese die Modellgüte verbessern (siehe Abschnitt 3). Bei Berücksichtigung von Unterschieden zwischen hydrologischen Großräumen steigt der Anteil der erklärten Varianz der (Box-Cox-transformierten) Nitratkonzentrationen von 21% auf 27% ($R^2_{adj} = 0,25$). Auch kategoriale erklärende Variablen wie der geochemische Gesteinstyp können mit Hilfe von Indikatorvariablen („Dummy-Variablen“) durch eine Reihe von 0/1-Werten repräsentiert werden.

Nichtlineare Beziehungen können systematischer durch nichtlineare Regressionsverfahren wie das additive Modell (*generalised additive model*, GAM) erfasst werden, welches es erlaubt, einzelne oder alle Terme $\beta_i f_i(\mathbf{x})$ durch Spline-Transformationen $s_i(f_i(\mathbf{x}))$ zu ersetzen (Wood, 2017); dies ist jedoch nur für quantitative (reellwertige) Variablen erforderlich. Dank der additiven Struktur ist das Modell vollständig visualisierbar und interpretierbar. Bei einer deutschlandweiten Modellierung erhöht dieses Vorgehen den R^2_{adj} jedoch nur um 0,018.

Variablenselektions- und Regularisierungsverfahren in linearen wie auch in additiven Modellen helfen, evidenzbasiert ein Trendmodell mit angemessener Komplexität zu identifizieren. Hierzu gehören das Akaike-Informationskriterium (AIC) und das Schwarz-Bayes-Informationskriterium (BIC) sowie Bestrafungsterme wie etwa in der LASSO-Methode (*least absolute shrinkage and selection operator*; vgl. James et al., 2021).

Weiterhin existieren geographisch gewichtete Regressionsmodelle (*geographically weighted regression*) für räumlich variierende Regressionskoeffizienten, $\beta_i(\mathbf{x})$ (Fotheringham et al., 2002). Die Modellanpassung dieser semiparametrischen Interaktionsterme ist jedoch fehlerträchtig.

Ohne Anspruch auf Vollständigkeit hinsichtlich weniger etablierter geostatistischer Verfahren erheben zu können, seien ferner geostatistische Copula-Modelle erwähnt, welche lokale Abhängigkeiten nicht-normalverteilter räumlicher Zufallsfelder repräsentieren und zuletzt auch um die Möglichkeit der Berücksichtigung von erklärenden Variablen erweitert wurden (Gnann et al., 2018; Bárdossy & Hörning, 2023). Diese Modelle sind jedoch im Vergleich zu den hier betrachteten Trendmodellen mathematisch und konzeptionell verhältnismäßig komplex, da Kreuz-Korrelogramme aller Variablen(paare) betrachtet werden müssen (Gnann et al., 2018).

Aufgrund ihrer größeren Flexibilität bei der Modellierung von nichtlinearen Beziehungen und Wechselwirkungen höherer Ordnung (d.h. zwischen mehreren erklärenden Variablen) haben verschiedene Modelle des maschinellen Lernens in den vergangenen Jahren an Beliebtheit bei der Modellierung des räumlichen Trends von Umweltvariablen gewonnen (z.B. Knoll et al., 2020; Petermann et al., 2021; Karimanzira et al., 2023), auch in Kombination mit der geostatistischen

Regionalisierung von Modellfehlern (Koch et al., 2019). So wird beim *random forest regression kriging* (RFRK) beispielsweise die Trendkomponente durch einen Random Forest (Breiman, 2001) modelliert und die Modellfehler durch OK interpoliert (Koch et al., 2019), und Sekulic et al. (2020) entwickelten eine räumliche Abwandlung dieses Verfahrens. Erste Experimente mit tiefen neuronalen Netzen (Deep Learning) wiesen keine Überlegenheit gegenüber Random Forest nach (Karimanzira et al., 2023), jedoch kann diese Vergleichsstudie aufgrund methodischer Einschränkungen keine abschließende Antwort liefern.

2.3 Vergleichende Bewertung von Regionalisierungsverfahren

Mit seinen Worten „*All models are wrong, [but] some are useful*“ brachte der Statistiker George Box bereits in den 1970er Jahren die Spannung zwischen der (vielleicht vergeblichen) Suche nach dem einen, „wahren“ Modell und der nutzungsorientierten Modellbewertung zum Ausdruck. Im Hinblick auf die vorliegende Fragestellung der Gebietsausweisung nitratbelasteter Gebiete bedeutet dies, dass über eine wünschenswerte Angemessenheit der statistischen Modellannahmen hinaus mehrere Qualitätsdimensionen zu berücksichtigen sind: 1. Prozessadäquatheit, 2. Nachvollziehbarkeit und Interpretierbarkeit, und 3. Vorhersagegüte.

2.3.1 Qualitative Verfahrensbewertung

Relevante Umweltprozesse sollten auf adäquate, d.h. auf möglichst plausible Weise durch das Modell zu repräsentieren, und zwar nicht nur in Form der verwendeten Daten, sondern auch hinsichtlich der Modellstruktur. Dies kann je nach Art des Modells auf unterschiedliche Weisen erzielt werden, beispielsweise in räumlichen linearen oder additiven Trendmodellen mit Hilfe von Interaktionstermen (siehe Abschnitt 2.2) und nichtstationären Abhängigkeitsstrukturen (siehe Abschnitt 2.6.1). Im maschinellen Lernen wird dagegen aktuell mit physikalisch informierten, hybriden Modellen experimentiert, welche hochflexiblen Modellen etwa des Deep Learning eine physikalisch oder konzeptionell plausible (Teil-) Struktur aufzwingen (Reichstein et al., 2019).

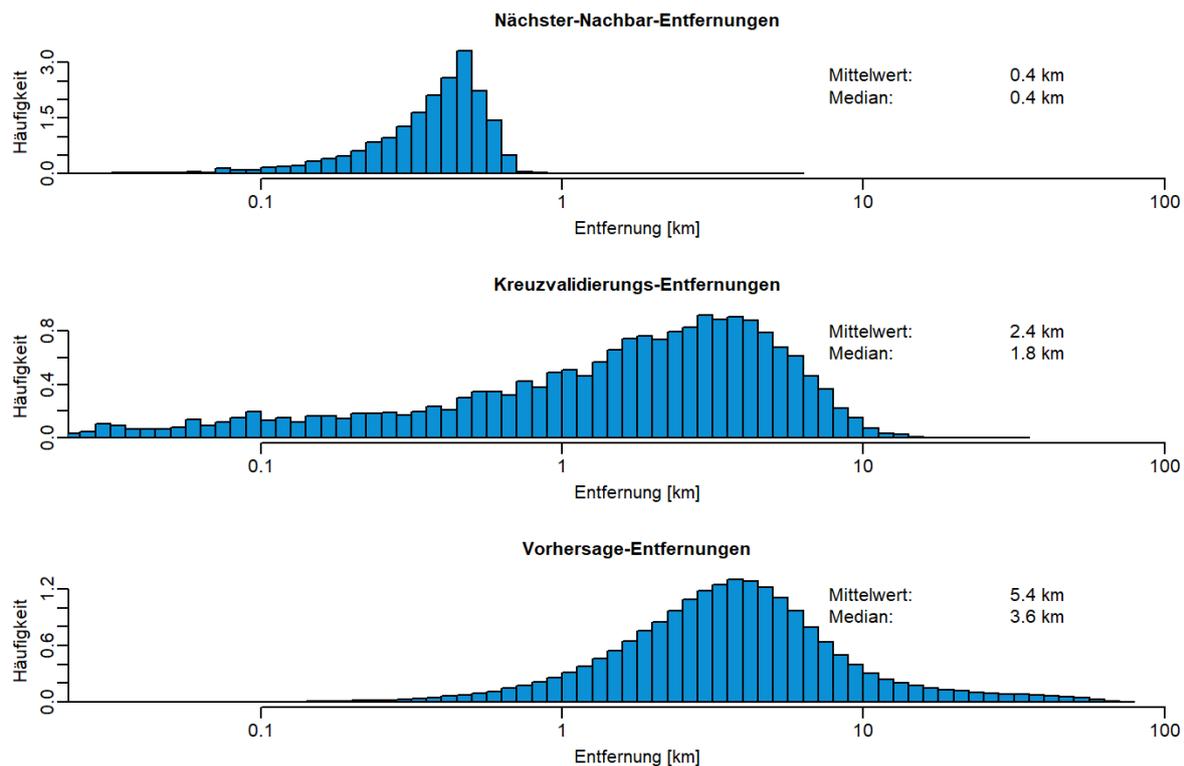
Um die konzeptionelle Plausibilität zu prüfen und generell der Nachvollziehbarkeit von Verwaltungsentscheidungen im Grundwasserschutz gerecht zu werden, sollten die verwendeten Modelle ferner interpretiert werden, also interpretierbar sein (vgl. Murdoch et al., 2019). Dies ist ein sehr starkes Argument für die Nutzung von („mindestens“) additiven Modellen (Rudin, 2019), einschließlich linearer Modelle, wie in Abschnitt 2.2 für Trendmodelle der Regionalisierung vorgestellt. Die Interpretierbarkeit von Black-Box-Modellen wie des Random Forest ist hingegen prinzipiell stark eingeschränkt (Rudin, 2019) und erschöpft sich in der Praxis in vereinfachten, methodisch eingeschränkten Verfahren wie Permutationswichtigkeit (Knoll et al. 2020; Koch et al., 2020) oder den ebenfalls recht begrenzten und nicht unproblematischen partiellen Abhängigkeitsdiagrammen (Karimanzira et al., 2023; vgl. Hooker et al., 2021; Murdoch et al., 2019; Molnar, 2023). Neben dem technischen Problem, dass diese Interpretationsverfahren fragwürdige Extrapolationen im Merkmalsraum vornehmen, wenn erklärende Variablen miteinander korreliert sind, ist hier das inhaltliche Problem hervorzuheben, dass komplexe Interaktionen, die ja gerade die Grundlage der untersuchten Modelle, durch eindimensionale Betrachtungen ignoriert werden (Molnar et al., 2023).

2.3.2 Quantitative Bewertung der Vorhersagegüte

Gegenüber den genannten qualitativen Kriterien rücken in vielen Modellvergleichen quantitative Maße der Vorhersagegüte in den Vordergrund, was mit Blick auf eine optimale Nutzung der vorhandenen Messwerte nachvollziehbar ist (z.B. Karimanzira et al., 2023; Ohlert et al., 2023a). Ein gängiges Fehlermaß ist hierbei die Wurzel des mittleren quadratischen Vorhersagefehlers

(*root mean-square error*, RMSE) der Regionalisierung. Dieses Fehlermaß wird häufig mit Hilfe einer **Kreuzvalidierung** (KV) geschätzt, d.h. mit einem rechnergestützten Schätzverfahren, bei dem nacheinander zufällig ausgewählte Teildatensätze für Validierungszwecke aus der Lernstichprobe des Modells entfernt werden (James et al., 2021). Ein Spezialfall ist die Leave-one-out-KV, bei der jeweils nur ein Datenpunkt aus dem Lerndatensatz entfernt wird, was jedoch zu geringer Variabilität in den trainierten Modellen führt.

Abbildung 3: Histogramme von Vorhersageentfernungen in Leave-one-out-Kreuzvalidierung, Zehnfach-Kreuzvalidierung und bei bundesweiter Regionalisierung



Man beachte die logarithmisch skalierte x-Achse! Nächster-Nachbar-Entfernungen wurden innerhalb des Messstellendatensatzes berechnet. Kreuzvalidierungs-Entfernungen sind minimale Entfernungen von Vorhersagepunkten zum Lerndatensatz in einer nicht-räumlichen Zehnfach-Kreuzvalidierung. Vorhersageentfernungen sind minimale Entfernungen von Punkten auf einem bundesweiten Vorhersage-Gitternetz mit 1 km × 1 km Auflösung zum Datensatz der Messstellen.

Quelle: eigene Darstellung, A. Brenning.

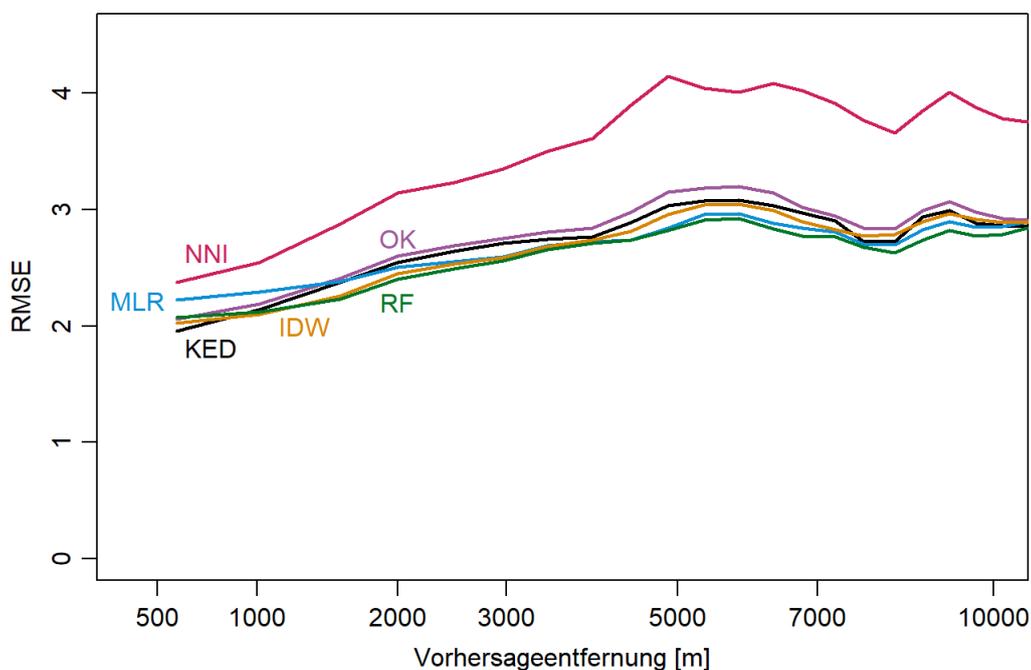
Diese Schätzverfahren können für räumliche Modelle ferner verzerrt sein, typischerweise überoptimistisch, sofern die räumlichen Entfernungen der Validierungspunkte von der Lernstichprobe nicht den in der Vorhersagesituation vorgefundenen Vorhersageentfernungen entsprechen (Brenning 2023). Um dies zu berücksichtigen, können räumliche KV-Verfahren eingesetzt werden (Brenning, 2012; Schratz et al., 2021; Milà et al., 2022) oder eine distanzbasierte Betrachtung angestrebt werden, die eine differenziertere Beurteilung der Vorhersagegüte über unterschiedliche Vorhersagedistanzen hinweg erlaubt (Brenning, 2023).

Vorhersagedistanzen in der bundesweiten Regionalisierung der Nitratbelastung unterscheiden sich sehr deutlich zwischen der Leave-one-out-KV, für die die Nächster-Nachbar-Entfernung ausschlaggebend ist, der nicht-räumlichen KV und der Vorhersagesituation im Zielgebiet (Abbildung 3). KV-Ergebnisse existierender bundesweiter Studien können daher verzerrt sein (z.B. Knoll et al., 2020; Karimanzira et al., 2023: nicht-räumliche Zehnfach-KV; Ohlert et al., 2023a, 2023b: Leave-one-out-KV).

Von kritischer Bedeutung ist ferner die Wahl der verwendeten **Fehlermaße**. Hier ist es zwingend erforderlich, sich auf die Nutzeranforderung einer Gebietsausweisung zu fokussieren, bei der es sich um ein „diagnostisches“, dichotomes Klassifikationsproblem handelt. Die in Vergleichen von Regionalisierungsverfahren – auch der Nitratbelastung des Grundwassers – bislang im Vordergrund stehenden Fehlermaße konzentrieren sich dagegen fast ausschließlich auf die Vorhersagefehler der Interpolation (vgl. Karimanzira et al., 2023); diese stehen jedoch in keiner direkten Beziehung zum eigentlichen Problem. So kann eine tatsächlich unbelastete Fläche mit sehr niedrigen Nitratkonzentrationen selbst dann korrekt als unbelastet klassifiziert werden, wenn der Vorhersagefehler an dieser Stelle groß ist, während an Stellen mit Nitratkonzentrationen nahe am Schwellenwert bereits ein kleiner Vorhersagefehler eine Missklassifikation nach sich ziehen kann. Zu diskutieren ist in diesem Zusammenhang auch die Frage, ob im Hinblick auf eine Ausweisung nitratbelasteter Gebiete eine irrtümliche Ausweisung (falsch-positive Klassifikation) gegenüber einer irrtümlichen Nichtausweisung (falsch-negative Klassifikation) gleich zu gewichten ist.

Unter Berücksichtigung dieser beiden Aspekte – der Wahl von Fehlermaß und Schätzverfahren – sind die derzeit vorliegenden quantitativen Methodenvergleiche äußerst kritisch zu betrachten und leider von geringer Aussagekraft (z.B. Ohlert et al., 2023a; Karimanzira et al., 2023).

Abbildung 4: Vorhersagefehlerprofile für NNI, IDW, OK, KED, MLR und RF



Verglichene Verfahren: NNI: Nächster-Nachbar-Interpolation („Voronoi-Verfahren“); IDW: Inverse Distanzgewichtung; OK: Ordinary Kriging; KED: Kriging mit externer Drift; MLR: multiple lineare Regression; RF: Random Forest. Die Berechnung erfolgt deutschlandweit nach dem Verfahren von Brenning (2023). Der Median der Vorhersageentfernung beträgt dabei 3,6 km. Die hier verwendeten Modelle nutzen teils vereinfachte Einstellungen zur Reduktion der Rechenzeit. Quelle: eigene Darstellung, A. Brenning.

Um diese Lücke zu füllen, wurden für ausgewählte Verfahren für das Pilotgebiet räumliche Vorhersagefehlerprofile berechnet, und zwar sowohl für den RMSE (Abbildung 4) als auch für die Über-/Unterausweisung nitratbelasteter Flächenanteile (Abbildung 5). Die Ergebnisse weisen darauf hin, dass die verwendeten Regionalisierungsverfahren IDW, OK, KED und RF trotz ihrer erheblichen mathematischen und konzeptionellen Unterschiede vergleichbare Regionalisierungsfähigkeiten aufweisen und lediglich die schwache NNI deutlich übertreffen (Abbildung 4).

Gleichzeitig ist bemerkenswert, dass diese Verfahren einen viel zu geringen Flächenanteil als nitratbelastet ausweisen (Abbildung 5). (Es sei angemerkt, dass dies *keine* direkten Rückschlüsse auf mögliche Verzerrungen in den aktuell existierenden Gebietsausweisungen erlaubt, da in der hier vorliegenden Analyse Box-Cox-transformierte Daten verwendet wurden; vgl. hierzu Abschnitt 2.6.2 sowie Tabelle 1.)

2.4 Modellierung von Überschreitungsflächen und -wahrscheinlichkeiten

Eigentliches Ziel ist es, innerhalb der nach §3 AVV GeA zu betrachtenden GWK immisionsbasiert diejenigen Teilgebiete

$$F := \{\mathbf{x} \in D: Z(\mathbf{x}) > \theta\}$$

abzugrenzen, die belastet sind (§5(1) AVV GeA). Grundsätzlich ist lediglich die Überschreitung des Schwellenwertes (bzw. ggf. Vorhandensein eines Trends) relevant, nicht jedoch die zu erwartende Höhe der Belastung selbst.

Die Gebietsausweisung kann somit auch als Klassifikations- oder Diagnoseproblem betrachtet werden. Während die in der Literatur fast ausschließlich betrachteten Interpolationsfehlermaße (z.B. RMSE) keine unmittelbare Relevanz für die diagnostische Güte haben, rücken somit andere Kriterien in den Vordergrund, die auf die Fähigkeit abzielen, tatsächlich (un-) belastete Gebiete auch als (un-) belastet auszuweisen. Interpolationsfehlermaße können hierbei in die Irre führen.

Als Referenzpunkt kann hierbei der Flächenanteil dienen, der tatsächlich nitratbelastet ist, also

$$f := |F|/|D|.$$

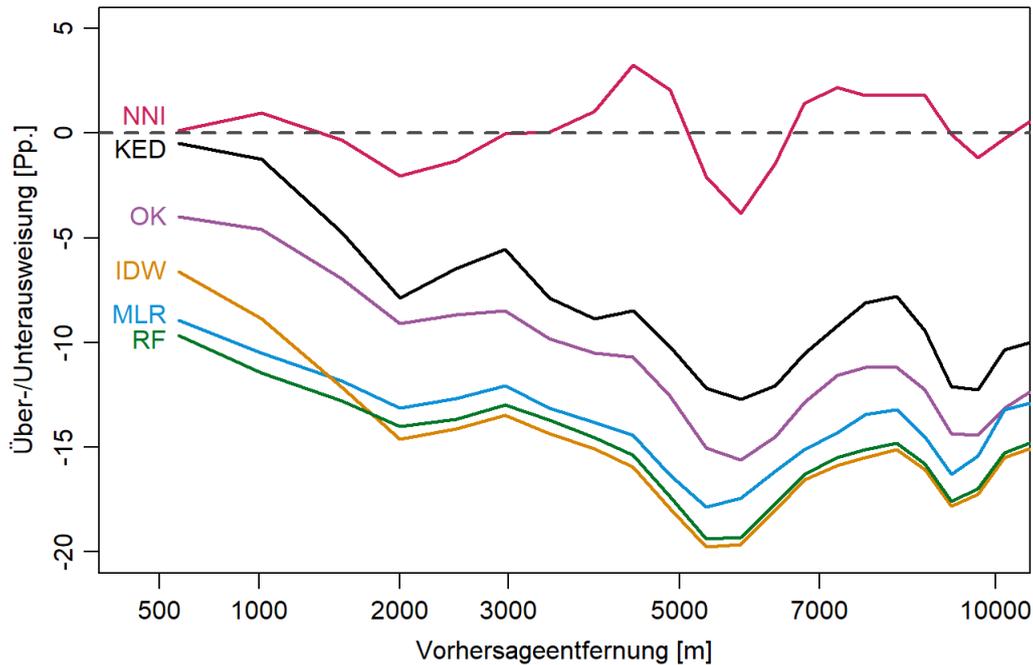
Dieser Flächenanteil ist zwar unbekannt, kann jedoch mit Hilfe eines im Folgenden beschriebenen geostatistischen Verfahrens geschätzt und zur Charakterisierung der Unsicherheit mit einem Konfidenzintervall versehen werden.

Gebietsausweisungen werden „klassisch“ schlichtweg durch Anwendung des Nitrat-Schwellenwertes $\theta = 50$ mg/l auf das Ergebnis der Regionalisierung $\{\hat{Z}(\mathbf{x}) : \mathbf{x} \in D\}$, durchgeführt,

$$\tilde{F} := \{\mathbf{x} \in D: \hat{Z}(\mathbf{x}) > \theta\}, \quad \tilde{f} := |\tilde{F}|/|D|.$$

Dies ist länderübergreifend die aktuelle Praxis. Hierbei bleibt jedoch unberücksichtigt, dass auch in vielen Gebieten, in denen das Interpolationsergebnis den Schwellenwert knapp unterschreitet, mit einer gewissen, nicht zu vernachlässigenden Wahrscheinlichkeit überschritten ist. Da die Messwerte überwiegend unter dem Schwellenwert liegen, wird dies nicht durch den gegenteiligen Effekt bei knapp überschrittenem Schwellenwert ausgeglichen. Aufgrund dieser Tatsache ist davon auszugehen, dass die so ausgewiesene Fläche $|\tilde{F}|$ die wahre Fläche deutlich unterschätzt. Dieses Phänomen ist seit Langem bekannt (z.B. Burrough et al., 2015) und vermeidbar, wie im Folgenden ausgeführt.

Abbildung 5: Vorhersagefehlerprofile für die Über-/Unterausweisung nitratbelasteter Flächen mit NNI, IDW, OK, KED, MLR und RF



Verglichene Verfahren: NNI: Nächster-Nachbar-Interpolation („Voronoi-Verfahren“); IDW: Inverse Distanzgewichtung; OK: Ordinary Kriging; KED: Kriging mit externer Drift; MLR: multiple lineare Regression; RF: Random Forest.

Die Berechnung erfolgt deutschlandweit nach dem Verfahren von Brenning (2023). Der Median der Vorhersageentfernung beträgt dabei 3,6 km. Die hier verwendeten Modelle nutzen teils vereinfachte Einstellungen zur Reduktion der Rechenzeit.

Quelle: eigene Darstellung, A. Brenning.

Um diese Verzerrung zu berücksichtigen und die wahre Überschreitungsfläche $|F|$ unverzerrt – d.h. ohne systematischen Fehler – zu bestimmen, ist es erforderlich, die Wahrscheinlichkeit

$$p(\mathbf{x}) = P(Z(\mathbf{x}) > \theta)$$

zu modellieren und durch $\hat{p}(\mathbf{x})$ zu schätzen bzw. räumlich vorherzusagen. Sofern $Z(\mathbf{x})$ normalverteilt ist und die Vorhersagevarianz $\sigma^2(\mathbf{x})$ bekannt ist, kann dies anhand der Perzentile der Standardnormalverteilung erfolgen:

$$\hat{p}(\mathbf{x}) := F_N^{-1} \left(\frac{\hat{Z}(\mathbf{x}) - \theta}{\sigma(\mathbf{x})} \right)$$

wobei F_N^{-1} die Umkehrfunktion der kumulativen Verteilungsfunktion der Standard-Normalverteilung ist. Im Falle von Kriging-Verfahren wie KED ist die Kriging-Varianz ein gültiger Schätzer für $\sigma^2(\mathbf{x})$, sofern die Annahmen des verwendeten geostatistischen Modells zutreffen. Der Umgang mit nicht normalverteilten Zielvariablen wie dem Nitratgehalt wird in Abschnitt 2.6.2 (und Abschnitt 2.6.3) behandelt.

Analog zu den Ausführungen zu Trendmodellen sind auch hier neben rein geostatistischen Regionalisierungsverfahren auch nichtlineare Modelle (z.B. additives Modell, Random Forest) und deren Kombination mit geostatistischer Charakterisierung der Vorhersageunsicherheiten denkbar (z.B. Koch et al., 2019). Ebenso lässt sich $\hat{p}(\mathbf{x})$ direkt mit Hilfe des geostatistischen Verfahrens des Indikator-Kriging (Cressie, 1993), mit logistischen linearen oder additiven Modellen oder mit Klassifikationsverfahren des maschinellen Lernens regionalisieren. Da dies eine Dichotomisierung der regionalisierten Werte $z(\mathbf{x}_i)$ in Indikatorwerte 0 und 1 für ein Unter- und

Überschreiten des Schwellenwertes erfordert, gehen jedoch relevante Informationen verloren und die Unsicherheiten der Vorhersage erhöhen sich. Diese Vorgehensweise wird hier daher nicht weiterverfolgt.

Tabelle 1: Bestimmung des Überschreitungsflächenanteils im Pilotgebiet mit NNI, IDW, OK, KED, RF und BGS

Verfahren	Untransformierte Daten	Logarithmisch transformiert	Box-Cox-transformiert
NNI	18,8%	18,8%	18,8%
IDW	13,5%	1,8%	3,4%
OK	15,7%	5,6%	7,3%
KED	26,7%	13,1%	16,5%
RF	16,0%	0,4%	1,2%
BGS ($\phi = 0,50$) „Standard-“ Kriterium	– *	– *	8,3%
BGS ($\phi = 0,35$) 5. Perzentil des Flächenanteils	– *	– *	18,1%
BGS ($\phi = 0,33$) unverzerrter Flächenanteil	– *	– *	20,1%

* Nicht ermittelt, da Verteilungsannahmen nicht erfüllt.

Verglichene Verfahren: NNI: Nächster-Nachbar-Interpolation („Voronoi-Verfahren“); IDW: Inverse Distanzgewichtung; OK: Ordinary Kriging; KED: Kriging mit externer Drift; MLR: multiple lineare Regression; RF: Random Forest; BGS: bedingte geostatistische Simulation.

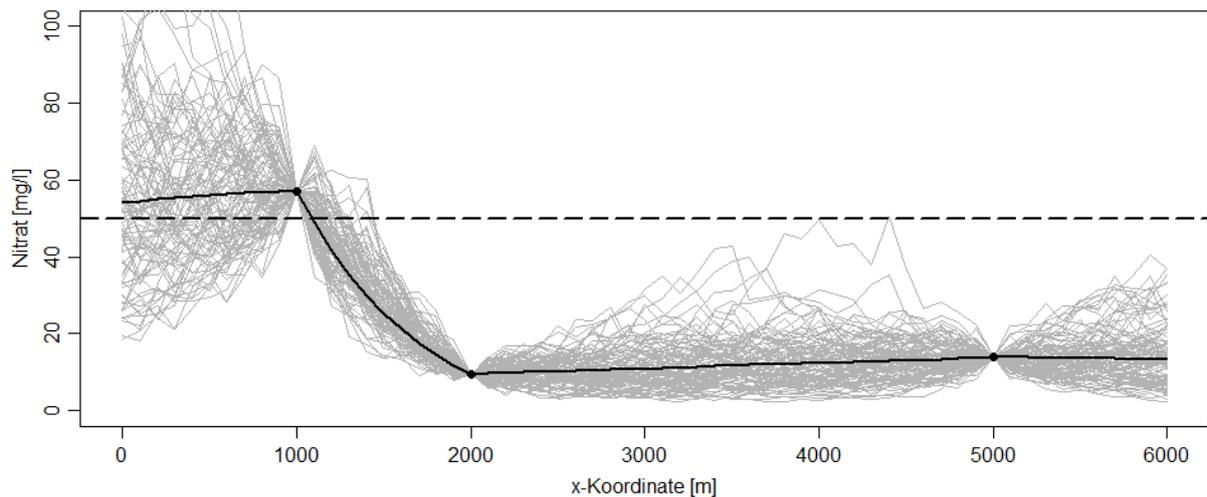
Sind die Annahmen des jeweiligen Modells erfüllt und die Schätzer $\hat{p}(\mathbf{x})$ unverzerrt, so lässt sich der Flächenanteil des Überschreitungsgebietes unverzerrt durch

$$\hat{f} := \frac{1}{|D|} \int_D \hat{p}(\mathbf{x}) \, d\mathbf{x}$$

schätzen, d.h. in der Praxis numerisch durch Mittelwertbildung über die auf einem Gitternetz vorhergesagten Wahrscheinlichkeiten.

Die **bedingte geostatistische Simulation** (BGS) ist eine alternative Herangehensweise, die die Möglichkeit eröffnet, die Schätzunsicherheit von \hat{f} zu bestimmen. In der Statistik versteht man unter Simulation die Erzeugung von zufälligen Datensätzen, die einer vorgegebenen Wahrscheinlichkeitsverteilung folgen, mit Hilfe von (Pseudo-) Zufallszahlengeneratoren. Bei einer BGS geschieht dies so, dass die gegebenen Bedingungen – hier: die beobachteten Messwerte an den Messstellen – eingehalten werden. Abbildung 6 illustriert dieses Verhalten exemplarisch anhand fiktiver Messwerte entlang eines Profils. Da diese Messwerte bei Vorliegen einer räumlichen Autokorrelation auch die Wahrscheinlichkeitsverteilung der unbeobachteten Nitratgehalte in ihrer Umgebung „einengen“, ist es ein besonderes Merkmal der BGS, dass diese Abhängigkeitsbeziehungen ebenso berücksichtigt werden wie die des zugrundeliegenden Trendmodells.

Abbildung 6: Ergebnisse bedingter geostatistischer Simulationen basierend auf Messungen an drei fiktiven Messstellen



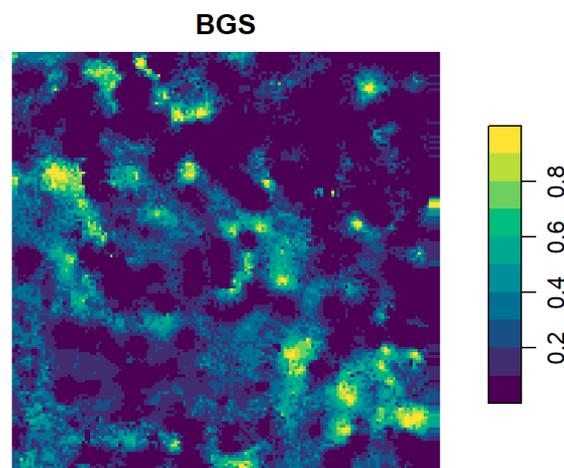
Das Diagramm zeigt 100 geostatistische Simulationen, die bedingt unter den drei (fiktiven) Messwerten (schwarze Punkte) erzeugt wurden. Die dickere schwarze Linie ist der Mittelwert über 1000 Simulationen und entspricht daher der Kriging-Vorhersage.

Quelle: eigene Darstellung, A. Brenning.

Insgesamt handelt es sich bei der BGS um ein zwar rechenintensives, aber durchaus etabliertes und erprobtes Verfahren (z.B. Borgoni et al., 2010 im Kontext der Radonbelastung und Beloconi et al., 2018 mit verwandten Bayes'schen Methoden im Kontext von Feinstaub). Abbildung 7 zeigt die per BGS ermittelten Überschreitungswahrscheinlichkeiten im Pilotgebiet.

Das Konfidenzintervall des Überschreitungflächenanteils (einseitig oder beidseitig) inklusive des unverzerrten Schätzers kann alternativ auch ohne BGS theoretisch hergeleitet werden. Hierzu benötigt es neben der geostatistischen Regionalisierung aber auch eines Schätzers aller paarweisen Kovarianzen des Zufallsfeldes bzw. seiner Diskretisierung auf D . BGS ist daher in der Praxis leichter implementierbar und informativer, da auch andere Merkmale und deren Variabilität abgeleitet werden können.

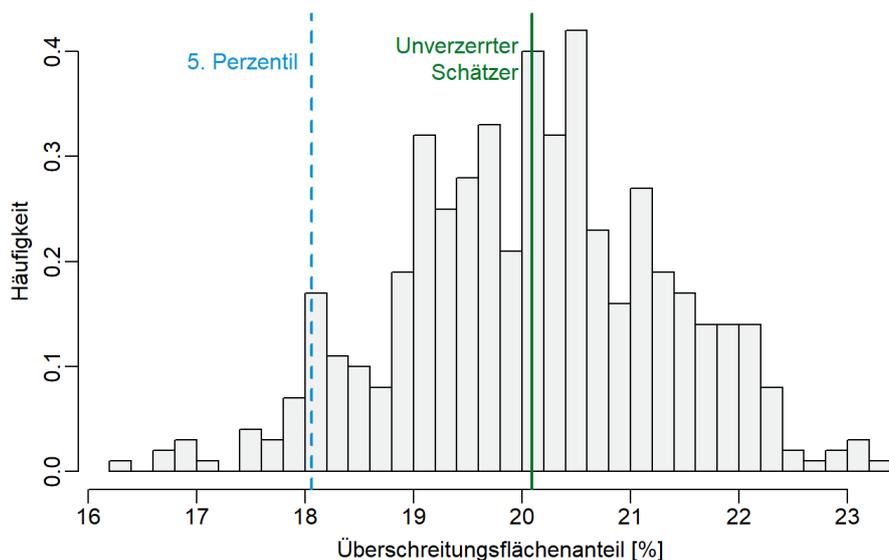
Abbildung 7: Regionalisierung der Überschreitungswahrscheinlichkeit für 50 mg/l im Pilotgebiet mit bedingter geostatistischer Simulation



Die berechneten Wahrscheinlichkeiten beruhen auf 500 bedingten geostatistischen Simulationen.

Quelle: eigene Darstellung, A. Brenning.

Abbildung 8: Bestimmung des Überschreitungsflächenanteils im Pilotgebiet und seine statistische Unsicherheit basierend auf bedingter geostatistischer Simulation



Das Histogramm zeigt das Ergebnis von 500 Simulationen. Eingezeichnet sind ferner der Mittelwert über alle Simulationen als unverzerrter Schätzer (20,1%) sowie das 5. Perzentil der Verteilung (18,1%) als Untergrenze eines (einseitigen) 95%-Konfidenzintervalls.

Quelle: eigene Darstellung, A. Brenning.

2.5 Wahl eines Schwellenwertes der Überschreitungswahrscheinlichkeit

Um auf der Grundlage der regionalisierten Überschreitungswahrscheinlichkeiten Teilgebiete als belastet auszuweisen, ist in einem weiteren Schritt die Wahl eines Schwellenwertes $0 < \phi < 1$ erforderlich. Mit dessen Hilfe werden Überschreitungsregionen bestimmt:

$$\hat{F}_\phi := \{\mathbf{x} \in D : \hat{p}(\mathbf{x}) > \phi\}$$

Der als belastet ausgewiesene Flächenanteil

$$\hat{f}_\phi := |\hat{F}_\phi|/|D|$$

einer Region nimmt zu, je niedriger ϕ gewählt wird. So wird bei einem sehr hohen Schwellenwert ϕ (z.B. 0,95) ein wesentlich kleinerer Flächenanteil als belastet ausgewiesen, als mit obigen Methoden geschätzt wurde (\hat{f} oder die Untergrenze von dessen Konfidenzintervall). Abbildung 9 zeigt diese Beziehung im Pilotgebiet anhand der BGS-Regionalisierung auf.

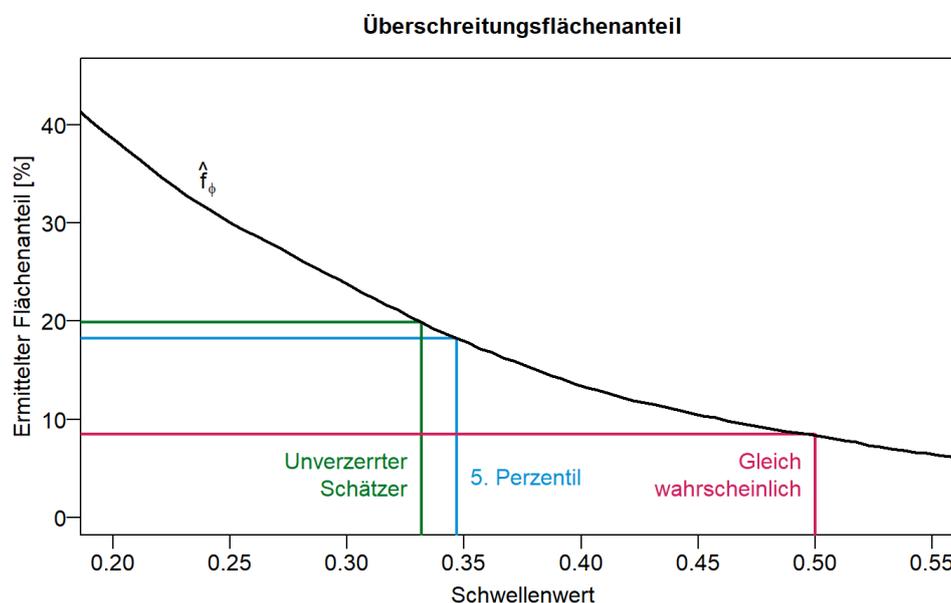
Gleichzeitig nimmt bei höherem ϕ der Anteil der tatsächlich belasteten Flächen ab, die korrekt als nitratbelastet identifiziert werden (niedrige Sensitivität des Diagnoseverfahrens). Umgekehrt nimmt mit niedrigerem ϕ der Anteil der tatsächlich belasteten Flächen zu, der richtigerweise als belastet ausgewiesen wird (hohe Sensitivität), und dafür steigt im Gegenzug der Anteil tatsächlich unbelasteter Flächen, die irrtümlich als belastet ausgewiesen werden (falsch-positive Diagnose).

Beide Fehlerarten verursachen „Kosten“: Während falsch-negative (fälschlich unterlassene) Gebietsausweisungen in ineffektivem Grundwasserschutz resultieren und somit gesundheitliche und volkswirtschaftliche Kosten verursachen können, resultieren falsch-positive (fälschlich durchgeführte) Ausweisungen nitratbelasteter Gebiete in Opportunitätskosten in landwirtschaftlichen Betrieben, deren Düngemittelapplikation eingeschränkt wird. Die AVV GeA enthält keine Hinweise darauf, in welcher Weise diese gegeneinander abzuwägen sind.

Unter Gesichtspunkten eines effektiven Grundwasserschutzes erscheint es grundsätzlich erstrebenswert, eine Gebietsausweisung anzustreben, deren Fläche sich am tatsächlich belasteten Flächenanteil orientiert, also an \hat{f} oder wenigstens der Untergrenze eines Konfidenzintervalls für \hat{f} (Abbildung 9), so dass mit (z.B.) 95% Sicherheit gewährleistet ist, dass die (wahre) Überschreitungsfläche mindestens so groß ist wie die ausgewiesenen Gebiete. In der Pilotregion resultieren diese beiden Ansätze in deutlich größeren Gebietsausweisungen als bei einem „herkömmlichen“ Vorgehen mit gängigen Interpolationsverfahren (Tabelle 1).

Abschließend sei darauf hingewiesen, dass sich die Modellierung von Überschreitungsregionen als ein eigenes Teilgebiet innerhalb der Geostatistik etabliert hat. Hier werden beispielsweise Methoden entwickelt, die eine Art „Konfidenzregion“ für die ermittelten Überschreitungsregionen bestimmen (Cressie & Suesse, 2020).

Abbildung 9: Abnahme des ermittelten Überschreitungsflächenanteils im Pilotgebiet mit zunehmendem Schwellenwert ϕ der Überschreitungswahrscheinlichkeit



Gekennzeichnet sind der mit bedingten geostatistischen Simulation unverzerrt geschätzte Überschreitungsflächenanteil (20,1%) sowie die Untergrenze eines (einseitigen) 95%-Konfidenzintervalls, 18,1%. „Gleich wahrscheinlich“ verwendet $\phi = 0,50$ als Entscheidungskriterium.

Quelle: eigene Darstellung, A. Brenning.

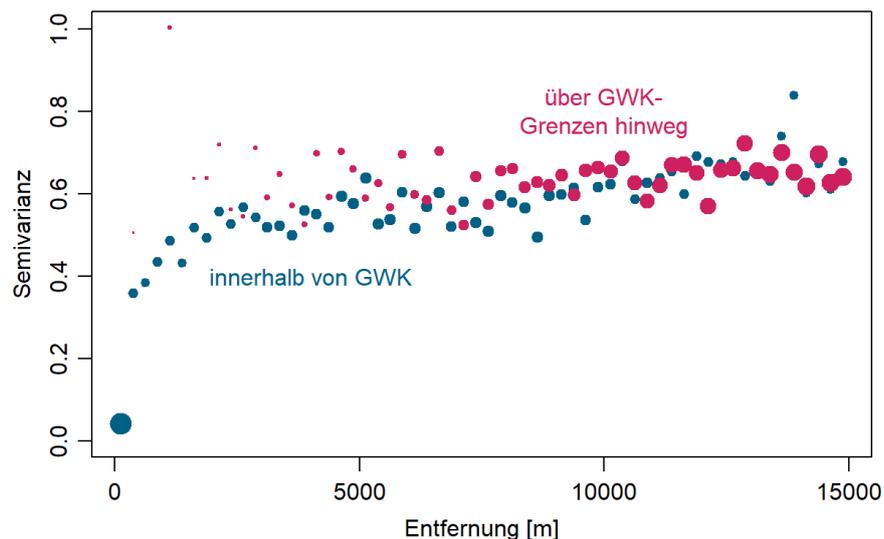
2.6 Spezifische Anforderungen und Herausforderungen

2.6.1 Nichtstationäre räumliche Abhängigkeit

Empirisch beobachtete räumliche Anhängigkeiten können verschiedene Ursachen haben, so etwa das Vorhandensein unbeobachteter Störgrößen, die nicht in das Trendmodell eingegangen sind, oder physikalische Transport- oder Durchmischungsprozesse, die die Ähnlichkeit von Stoffkonzentrationen über gewisse Entfernungen erhöhen. Im Kontext der Nitratbelastung sind beide Aspekte vorstellbar, da beispielsweise hydrogeologische Verhältnisse nur ungenau charakterisiert werden können (z.B. grobe Klassen von k_f -Werten), und lateraler Stofftransport sowie zeitliche Mittelwertbildung statistisch in einer Mischungsverteilung resultieren, jedenfalls soweit keine hydraulische Barriere vorliegt.

Hieraus resultieren zwei Arten von räumlichen Autokorrelationen: (1) Innerhalb von GWK, welche gemäß §5 (1) AVV GeA bei der Regionalisierung zwingend zu berücksichtigen sind, und (2) über GWK-Grenzen hinweg. Abbildung 10 zeigt, dass die Stärke von räumlichen Autokorrelationen innerhalb desselben GWK ähnlich groß ist wie über GWK-Grenzen hinweg. Vor diesem Hintergrund erscheint es nicht sinnvoll, an GWK-Grenzen eine „harte“ Dekorrelation zu modellieren, welche auch nicht durch die AVV GeA gefordert wird.

Abbildung 10: Semivariogramme der Regressionsresiduen innerhalb über GWK-Grenzen hinweg sowie innerhalb desselben GWK

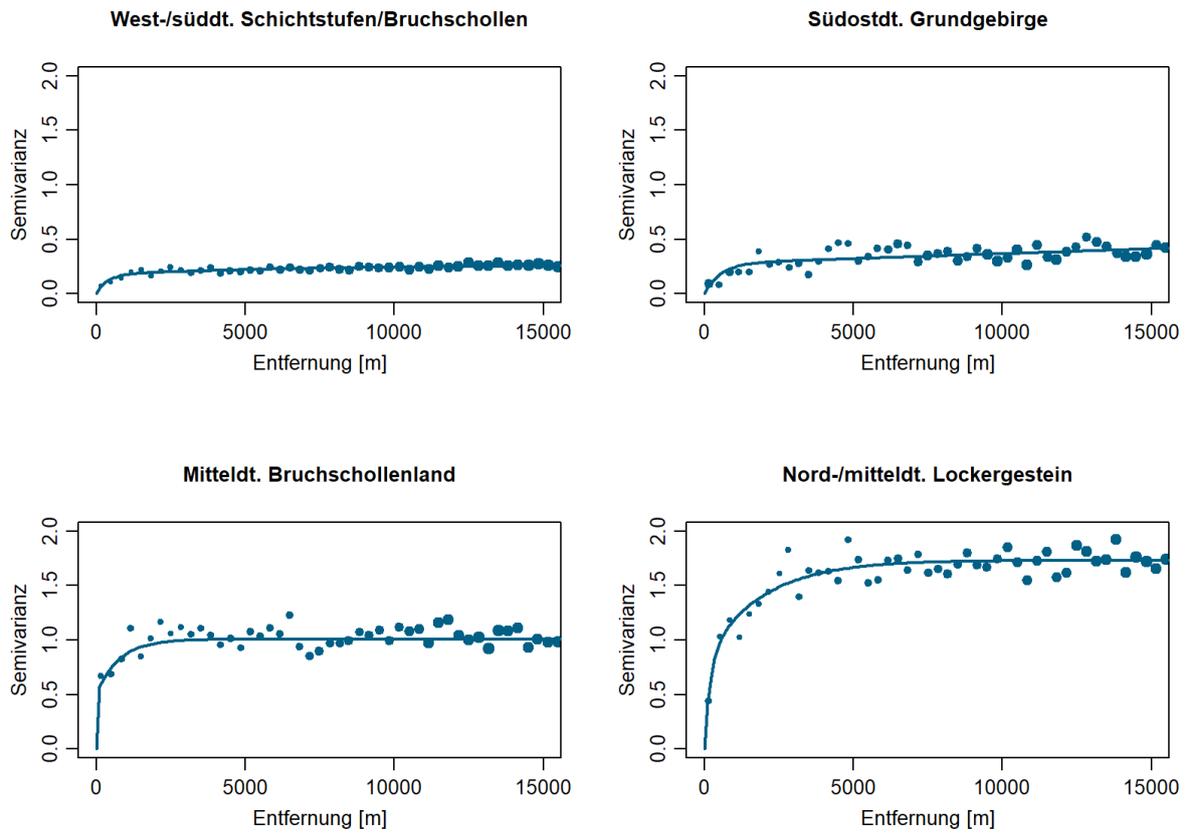


Die Semivariogramme wurden für die studentisierten Residuen von Regressionsmodellen für Box-Cox-transformierte Daten berechnet. Punktflächen sind proportional zur Anzahl verwendeter Punktepaare. Da GWK-Grenzbereiche bundesweit nur geringe Flächen einnehmen, liegen nur wenige Punktepaare für eine Semivarianz-Schätzung auf kurzen Distanzen über GWK-Grenzen hinweg zur Verfügung.

Quelle: eigene Darstellung, A. Brenning.

Räumliche Abhängigkeiten können ferner innerhalb unterschiedlicher hydrogeologischer oder geökologischer Teilräume unterschiedlich strukturiert sein. Dies wird vor allem bei Betrachtung von Semivariogrammen (der Residuen des Trendmodells) innerhalb hydrogeologischer Großräume deutlich (Abbildung 11). Dies ist der Nachweis einer nichtstationären räumlichen Abhängigkeit, welche durch geeignete Semivariogramm-Modelle berücksichtigt werden sollte. Dies ist von großer Bedeutung, da offenbar (nicht unerwartet) in hydrogeologisch homogeneren Großräumen mit weiträumigen, flach geneigten Porengrundwasserleitern größere Autokorrelationsweiten vorliegen, die eine bessere Interpolation von Punktmessungen in die Fläche erlauben. Eine finale Entscheidung über die Art der Parametrisierung dieser Nichtstationarität kann im Rahmen der Modellbildung auf der Grundlage von AIC oder BIC erfolgen bzw. durch diese Kriterien unterstützt werden. Ähnlich wie bei der Trendmodellierung deutet dies darauf hin, dass es sinnvoll ist, hydrogeologische Großräume getrennt zu modellieren.

Abbildung 11: Semivariogramme der Regressionsresiduen innerhalb ausgewählter hydrogeologischer Großräume



Die Semivariogramme wurden für die studentisierten Residuen von Regressionsmodellen für Box-Cox-transformierte Daten berechnet. Punktflächen in den empirischen Semivariogrammen sind proportional zur Anzahl Punktpaare, die in die Berechnung eingegangen sind.

Quelle: eigene Darstellung, A. Brenning.

2.6.2 Nicht-Normalverteilung

Eine verbreitete Sorge bei der Analyse und Regionalisierung von Nitratgehalten im Grundwasser und anderer Umweltvariablen ist das Vorliegen einer rechtsschiefen Verteilung der Daten, welche teils zum Anlass für eine logarithmische Transformation der Messwerte genommen wird (z.B. Wriedt et al., 2019). In diesem Zusammenhang ist jedoch zunächst festzustellen, dass die Unverzerrtheit der Vorhersagen beispielsweise des Kriging keine Normalverteilung erfordert. Ganz im Gegenteil kann die „naive“ Rücktransformation von Kriging- oder auch IDW-basierten Interpolationen logarithmierter Daten zu systematischen Verzerrungen der Regionalisierung auf der Originalskala führen (vgl. z.B. Cressie, 1993). Verzerrungskorrekturen existieren, sind jedoch gegenüber Verteilungsannahmen sensitiv (Cressie, 1993) und rücken bei einer Fokussierung auf die Berechnung von Überschreitungswahrscheinlichkeiten für einen Schwellenwert in den Hintergrund.

(Normal-) Verteilungsannahmen sind jedoch erforderlich, wenn mit Hilfe von Kriging oder der Kombination von Modellen des maschinellen Lernens mit der Geostatistik Vorhersageintervalle oder Überschreitungswahrscheinlichkeiten berechnet werden oder Unsicherheiten propagiert werden sollen. Bei Modellen mit Trendkomponente ist hierbei nicht die Verteilung der Messdaten selbst entscheidend (wie bspw. von Koch et al., 2019 angenommen), sondern die der Modellresiduen, für welche die Verteilungsannahmen getroffen werden.

Eine Annäherung an eine Normalverteilung kann mit Hilfe verschiedener Transformationen erreicht werden, wobei insbesondere die logarithmische Transformation und die Box-Cox-Transformation weit verbreitet sind. Bei der Box-Cox-Transformation handelt es sich um eine parametrische Transformation

$$Z^{(\lambda)}(\mathbf{x}) := \begin{cases} \frac{Z(\mathbf{x})^\lambda}{\lambda} & \text{für } \lambda \neq 0 \\ \ln(Z(\mathbf{x})) & \text{für } \lambda = 0 \end{cases}$$

mit einem reellwertigen Parameter λ , der empirisch bestimmt wird. Für $\lambda = 0$ erhält man als Spezialfall die logarithmische Transformation.

Ein aufwendigerer Ansatz ist die Verwendung der Normal-Score-Transformation (oder Gauß-Anamorphose), bei der die Verteilung der beobachteten Werte empirisch auf eine Normalverteilung abgebildet wird (z.B. Cressie, 1993; Wackernagel, 2003, Kap. 33).

Die Residuen eines deutschlandweiten linearen Modells der Nitratgehalte sind deutlich rechtsschief. Alle genannten Transformationen der Nitratgehalte resultieren in Residuenverteilungen, die näher an einer Normalverteilung sind. Die Box-Cox-Transformation (mit $\lambda = 0,2222$) erscheint für den vorliegenden Anwendungskontext angemessen und am praktikabelsten.

Kriging transformierter Daten wird als **trans-Gaußsches Kriging** bezeichnet (Cressie, 1993).

2.6.3 Bestimmungsgrenze

Messwerte des Nitratgehalts liegen bundesweit an etwa 6% der Messstellen unter der Bestimmungsgrenze von typischerweise 0,1 mg/l. Die Grundwasserverordnung sieht für diesen Fall eine Substitution durch die halbe Bestimmungsgrenze vor. Aus statistischer Sicht ist dieses Vorgehen nicht optimal, und es existieren angemessenere Ansätze (Helsel, 2005).

Durchgeführte Sensitivitätsanalysen zeigen jedoch, dass diese Problematik bei der Bestimmung von Überschreitungsgebieten keine substanziellen Auswirkungen hat. Dies ist darauf zurückzuführen, dass die Bestimmungsgrenze zwei Größenordnungen unter den maßgeblichen Schwellenwerten liegt und nur einen geringen Anteil der Messstellen betrifft.

Nichtsdestotrotz kann die aktuelle Regelung der Grundwasserverordnung bei anderen Substanzen mit häufigerer Unterschreitung der Bestimmungsgrenze und Schwellenwerten nahe an der Bestimmungsgrenze zu erheblichen Verzerrungen führen.

2.6.4 Unsicherheiten in erklärenden Variablen

Unsicherheiten in erklärenden Variablen sind ein Aspekt, der in der Regionalisierung und räumlichen Modellierung nur selten Beachtung findet, jedoch einen bedeutenden Beitrag zur Vorhersageunsicherheit leisten kann, gerade wenn stark generalisierte Variablen (z.B. hydrogeologische Charakteristika im Maßstab 1:500.000 oder mit 1 km Auflösung) mit Punktmessungen kombiniert und für eine Regionalisierung in höherer räumlicher Auflösung genutzt werden.

Sofern Fehler in erklärenden Variablen räumlich homogen sind, erscheint eine besondere Betrachtung auch nicht notwendig. Hervorzuheben ist jedoch, dass bei kategorialen Variablen, die eine Raumgliederung repräsentieren, entlang der Teilraumgrenzen Streifen mit erhöhter Unsicherheit entstehen. Beispielsweise ist bei der Abgrenzung hydrogeologischer Teilräume (Maßstab 1:500.000) mit Bereichen von mehreren Hundert Metern Breite zu rechnen, in denen die Zuordnung unsicher ist; an den Grenzen hydrogeologischer Großräume (1:250.000) reduzieren sich diese Bereiche schätzungsweise auf ± 250 m.

Eine pragmatische Vorgehensweise bei der Gebietsausweisung kann darin bestehen, in diesen Bereichen die Zugehörigkeit zur jeweiligen Klasse (Indikatorvariable mit Werten 0 und 1) durch einen linearen Übergang mit Werten von 0 bis 1 zu ersetzen (Fuzzy-Zugehörigkeit; vgl. De Smith et al., 2007). Denkbar ist auch eine probabilistische Vorgehensweise in Verbindung mit geostatistischer Simulation für die Bestimmung von Überschreitungswahrscheinlichkeiten.

2.6.5 Skalenwechsel

Zwar liegt auf der Hand, dass jegliche Regionalisierung von Nitratgehalt oder Überschreitungswahrscheinlichkeit in der Praxis auf einem Zielraster stattfindet und/oder auf einen Zielmaßstab abzielt, jedoch findet dies in den bislang vorgestellten Ansätzen keine Beachtung. Die AVV GeA schreibt eine Ausweisung „mindestens im Maßstab 1 : 25 000 oder in einem flächenscharfen digitalen System“ vor (§7 (2) AVV GeA), was einer Auflösung von etwa 12,5 m entspräche (Tabelle 2) und in keinem angemessenen Verhältnis zu Auflösung und Detaillierungsgrad relevanter Umweltzustandsvariablen steht (vgl. Abschnitt 3). Mit Blick auf diese Daten (siehe Abschnitt 3) erscheinen Zielauflösungen im Bereich zwischen 125 m und 1 km vertretbar (vgl. auch Koch et al., 2019, 100 m und Knoll et al., 2020, 1 km Auflösung mit vergleichbaren Eingangsdaten).

Tabelle 2: Gegenüberstellung von Kartenmaßstab, kleinster kartierbarer Einheit und empfohlener Raster-Auflösung für Geodaten

Kartenmaßstab	Kleinste kartierbare Einheit (Durchmesser in m)	Auflösung (m)
1:10.000	10	5
1:50.000	50	25
1:100.000	100	50
1:250.000	250	125
1:500.000	500	250
1:1.000.000	1000	500

Quelle: Nach ESRI, <https://www.esri.com/arcgis-blog/products/product/imagery/on-map-scale-and-raster-resolution/>, ursprünglich bei Tobler (1987).

Festzuhalten ist jedoch, dass der Skalenwechsel von Punktmessungen zu Flächen (z.B. Rasterzellen) mit einer räumlichen Aggregation (Mittelung) einhergeht. In der Geostatistik wird dies als Trägerwechsel (*change of support*) bezeichnet, und die Zielfläche als Block, B , beispielsweise eine Rasterzelle oder eine Referenzparzelle. An die Stelle von Punktvorhersagen $\hat{Z}(\mathbf{x}_0)$ tritt nun die Block-Vorhersage $\hat{Z}(B)$. Diese kann geostatistisch durch das so genannte **Block-Kriging** (mit Optimalitätseigenschaften analog zu OK und UK) bestimmt werden. Die Unsicherheiten dieser Vorhersage werden nicht durch das klassische Punkt-Semivariogramm, sondern durch das Punkt-zu-Block-Semivariogramm

$$\gamma(\mathbf{x}, B) = \frac{1}{|B|} \int_B \gamma(\mathbf{x}, \mathbf{u}) \, d\mathbf{u}$$

sowie die Block-zu-Block-Semivariogramme bestimmt (Cressie, 1993), wodurch sich die Vorhersageunsicherheiten mit zunehmender Größe der Blöcke (Rasterzellen) verringern.

Geostatistisch motiviert sind ferner die mit dem Block-Kriging verwandten Ansätze von Boumans et al. (2008) und Schafmeister et al. (2023). Es sind keine äquivalenten Skalenwechsel-Methoden im maschinellen Lernen bekannt, jedoch wäre grundsätzlich eine numerische Aggregation von höher aufgelösten Punkt-Vorhersagen auf die Ziel-Rasterzellen denkbar.

Es ist anzumerken, dass Block-Vorhersagen (selbstverständlich) nicht die Erwartung erfüllen (können), dass sie dem an einer Messstelle innerhalb des Blocks gemessenen Wert exakt übereinstimmen. In dieser Hinsicht ist die Anforderung der AVV GeA, dass um belastete Nitratmessstellen (Punktmessungen) zwingend eine belastete Fläche auszuweisen ist (§6 AVV GeA), zumindest aus statistischer Sicht kritisch zu hinterfragen, und zwar unabhängig vom verwendeten empirischen Ausweisungsverfahren. Aus praktischer Sicht könnte der AVV GeA diesbezüglich jedoch dadurch genüge getan werden, dass entsprechende Rasterzellen unabhängig vom Regionalisierungsergebnis das Attribut „belastet“ zugewiesen bekommen.

2.6.6 Trendbestimmung bei subkritischen Nitratgehalten

Ergänzend zu Gebieten mit Nitratkonzentrationen über 50 mg/l sind laut AVV GeA ergänzend jene Gebiete mit 37,5 bis 50 mg/l und einem steigenden Trend als nitratbelastet auszuweisen. Bundesweit fallen rund 8% aller hier betrachteten Messstellen in dieses Intervall, und bei 9% dieser subkritischen Messstellen wurde ein steigender Trend gemäß des in der Grundwasserverordnung festgeschriebenen Verfahrens festgestellt. Zentrale Aspekte der geltenden Regelung und der in ihrer Anlage 6 formulierten Vorgehensweise sind dabei aus statistischer Sicht kritisch zu hinterfragen (Brenning & Süße, 2023).

Zwar ist dieses Kriterium lediglich untergeordnet maßgeblich für eine Gebietsausweisung, es stellt sich dennoch die Frage nach der Regionalisierung von Trendschätzungen und ihrer Kombination mit Ergebnissen der Regionalisierung von Nitratkonzentrationen bzw. ihrer Überschreitungswahrscheinlichkeiten.

Ein möglicher Ansatz für die Regionalisierung von an Messstellen durchgeführten Trendtests besteht darin, die betreffende Teststatistik als räumliches Zufallsfeld zu betrachten, welches mit Hilfe geostatistischer Verfahren (OK, KED) regionalisiert wird. Auf diese Weise lässt sich – ähnlich der Regionalisierung von $\hat{p}(\mathbf{x})$ in Abschnitt 2.4 – die Wahrscheinlichkeit $\hat{p}^{(t)}(\mathbf{x})$ eines positiven Trends ermitteln. Ferner bezeichnet $\hat{p}^{(37,5-50)}(\mathbf{x})$ die vorhergesagte Wahrscheinlichkeit, dass die Nitratkonzentration zwischen 37,5 und 50 mg/l liegt.

Hieraus errechnet sich die Wahrscheinlichkeit, dass wenigstens eine der beiden Bedingungen erfüllt ist, aus

$$\hat{p}(\mathbf{x}) + \hat{p}^{(37,5-50)}(\mathbf{x}) \cdot \hat{p}^{(t)}(\mathbf{x}).$$

Aufgrund der Erfordernis (laut Grundwasserverordnung) eines signifikanten Trendtests wäre diese Kombination jedoch nur bei entsprechend hoher Wahrscheinlichkeit $\hat{p}^{(t)}(\mathbf{x})$ anzuwenden.

3 Datengrundlagen

Im Rahmen einer Regionalisierung können laut AVV GeA Anlage 2 „Hydrogeologie und Landbedeckung [...] als Zusatzkriterien ausgewertet werden“, was hier so verstanden wird, dass jegliche Merkmale, die in unmittelbarem Zusammenhang mit Grundwasserprozessen, Beschaffenheit des Grundwasserleiters sowie Landbedeckung stehen, grundsätzlich in Betracht kommen (siehe auch AVV GeA Anlage 4). Einschränkend dürfen nur Umweltzustandsdaten herangezogen werden, nicht jedoch landwirtschaftliche Emissionsdaten (UBA, 2023) oder Variablen, die die Denitrifikationskapazität beschreiben (z.B. Redox-Potenzial: Knoll et al., 2020).

Es werden drei übergeordnete Typen von Datengrundlagen unterschieden. Erstens wird eine allgemeine geoökologische Raumgliederung betrachtet, welche als Grundlage für eine prozessadäquate Identifikation homogener(er) Teilräume dienen kann. Neben hydrogeologischen Faktoren werden hier edaphische, klimatische und geomorphologische Aspekte sowie die Landbedeckung herangezogen (Abschnitt 3.2). Zweitens werden Datengrundlagen für prozessbezogene Umweltvariablen ermittelt, welche die lokalen hydrogeologischen, hydraulischen und hydrologischen Parameter abbilden sollen (Abschnitt 3.3). Teils leiten sich diese aus Grundwasser-Prozessmodellen ab. Von essenzieller Bedeutung ist jedoch, dass die Grundwasserbeschaffenheit aufgrund des lateralen Flusses eines Standorts nicht unmittelbar und kausal mit den lokalen Standortbedingungen an der Messstelle selbst in Zusammenhang steht. Besondere Bedeutung kommt vielmehr den Bedingungen im Anstrombereich der Messstelle zu. Aus diesem Grunde wird drittens die prozessbezogene Merkmalsextraktion diskutiert (Abschnitt 3.4).

Im Folgenden werden mögliche Datengrundlagen erfasst und hinsichtlich ihrer grundsätzlichen Eignung als Zusatzinformationen für die Regionalisierung vorläufig bewertet. Hierzu wird ermittelt, welchen Anteil der Varianz der Nitratverteilung deutschlandweit sowie innerhalb hydrogeologischer Großräume durch eine Variable erklärt wird; um die Anzahl von Klassen bei kategorialen Variablen zu berücksichtigen, wird hierzu der korrigierte R^2 (R^2_{adj}) verwendet.

Tabelle 3, Tabelle 4 und Tabelle 5 geben eine Übersicht über die im folgenden erläuterten Datengrundlagen. Zur besseren Einordnung der unterschiedlichen kartographischen Maßstäbe bzw. räumlichen Auflösungen wird auf Tabelle 2 verwiesen.

Tabelle 3: Datengrundlagen: 1. Geoökologische Raumgliederung

Variable	Maßstab / Auflösung	Merkmalstyp	Zusammenhang (R^2_{adj})	Nutzen der Interaktion mit HYRAUM
COP-DEM Höhe	25 m	Quantitativ	0,008	+0,018
LBM Landbedeckung	Mindestkartierfläche 1 ha	Kategorial: 7	0,023	+0,042
HYRAUM Großräume	1:500.000	Kategorial: 10	0,111	-
Boden-Klima-Räume	Gemeinde	Kategorial: 50	0,158	+0,062
UFZ Agrar-Umwelt-Räume	1 km	Kategorial: 16	0,091	+0,042

Als Maß für die Stärke des Zusammenhangs dient der korrigierte R^2 (R^2_{adj}). Der Nutzen der Interaktion mit HYRAUM beschreibt die Steigerung des R^2_{adj} durch Verwendung des Interaktionsterms.

3.1 Messstellen

Für die Auswertungen im Rahmen dieser Begutachtung werden Nitratmessstellen von Grundwassermessstellen in ganz Deutschland herangezogen, die auch für die Gebietsausweisung nach AVV GeA zur Verfügung stehen. Aufgrund der Kombination mit räumlichen Daten von erklärenden Variablen aus unterschiedlichen Quellen werden hier von den insgesamt rund 13.500 Messstellen lediglich 9844 Messstellen berücksichtigt. Beispielsweise wurden Messstellen ausgeschlossen, deren Landbedeckung als Wasserfläche klassifiziert wurde, oder an denen Informationen zum Humusgehalt oder der nFK fehlten. Im Rahmen einer operationellen Gebietsausweisung wären hier umfangreichere Datenvorprozessierungen einschließlich des Füllens von Datenlücken (Imputation) erforderlich. Aufgrund des hierfür erforderlichen Zeitbedarfs sowie der Fülle noch verbleibender Daten wird hier nur dieser reduzierte Datensatz analysiert.

Regionalisierungen werden hier ferner lediglich exemplarisch für eine Pilotregion von 150 km × 150 km Ausdehnung durchgeführt. In dieser Region stehen vollständige Daten von 471 Messstellen zur Verfügung.

3.2 Allgemeine geoökologische Raumgliederung

Für eine allgemeine geoökologische Raumgliederung eignen sich insbesondere die folgenden Datengrundlagen. Diese repräsentieren teils einzelne thematische Aspekte, teils eine Kombination mehrerer Dimensionen, und liegen in unterschiedlichen Auflösungen und Detaillierungsgraden vor.

Relief: Das Relief ist ein lokaler Proxy für Muster in einer Vielzahl potenziell relevanter Umweltzustandsvariablen, von (Topo-) Klima bis Boden- und Sedimentstrukturen. Neben der Geländehöhe selbst sind geomorphometrische Variablen wie das topographische Einzugsgebiet oder die mittlere Hangneigung im topographischen Einzugsgebiet potenziell relevant (siehe auch Abschnitt 3.4). Digitale Geländemodelle (DGM) repräsentieren die Geländeoberfläche (ohne Vegetation und Gebäude) und unterscheiden sich somit von Oberflächenmodellen (DOM). Mit dem Copernicus DEM (COP-DEM GLO-30) der European Space Agency liegt ein DOM in 1 Bogensekunde (etwa 25 m) Auflösung frei verfügbar in geeigneter Auflösung vor (AIRBUS, 2020). Für hydrologische Zwecke angemessener ist das hiervon abgeleitete, um Vegetation und Gebäude bereinigte FABDEM (Forest And Buildings Removed Copernicus DEM), welches unter nichtkommerzieller Lizenz frei verfügbar ist (Hawker & Neal, 2021; Hawker et al., 2022). Hier wurde jedoch lediglich das Copernicus DEM evaluiert.

Landbedeckung: An der Schnittstelle von Atmosphäre und Boden spielt die Landbedeckung eine zentrale Rolle im Wasser- und Stickstoffkreislauf, woraus sich ihre potenzielle Relevanz für die Nitratverteilung im Grundwasser ableitet, und zwar auch unabhängig von der tatsächlich vorliegenden Landnutzung und damit einhergehenden Emissionen. In Europa liegt mit dem CORINE Land Cover (CLC) Datensatz sowie in Deutschland mit dem Landbedeckungsmodell (LBM-DE) des BKG eine regelmäßig aktualisierte Datengrundlage vor.

Hydrogeologie: Prozessbestimmende hydrogeologische Merkmale folgen auf regionaler Ebene allgemeinen Mustern des geologischen Substrats und seiner inneren wie auch geomorphologischen Struktur, etwa in Gestalt großräumiger triassischer Poren-Grundwasserleiter in Schichtstufenlandschaften oder von Kluftgrundwasserleitern im Grundgebirge. Mit HYRAUM der BGR (2023) liegt eine dreistufige hierarchische hydrogeologische Raumgliederung im Maßstab 1:500.000 in 10, 36 und 247 Einheiten vor, wobei letztere aufgrund ihrer Kleingliedrigkeit nicht als erklärende Variable geeignet ist. HYRAUM wird als Proxy für eine Kombination hydrogeologischer Merkmale und für eine mögliche Unterteilung in Modelldomänen angesehen (siehe Abschnitte 2.2 und 2.6.1).

Boden-Klima-Räume: Boden- und Klimaeigenschaften spielen eine wichtige Rolle als Steuerungsfaktoren für die Grundwasserneubildung (GWN) und verdienen somit Beachtung als potenziell relevante Raumgliederung. Die am Julius-Kühn-Institut entwickelte Klassifikation Deutschlands in Boden-Klima-Räume (BKR; Roßberg et al., 2007) identifiziert Gebiete mit relativ homogenen Standortbedingungen hinsichtlich verschiedener Einflussfaktoren für die landwirtschaftliche Produktion mit Fokus auf Bodengüte und Klima. Die 50 BKR sind auf Gemeindeebene ausgewiesen. Ein anderer methodischer Ansatz wurde am UFZ bei der Abgrenzung von Agrar-Umwelt-Räumen (*archetypes of agri-environmental potential*, AEA) verfolgt (Beckmann et al., 2022). Diese Raumgliederung wendet ein datengetriebenes unüberwachtes Klassifikationsverfahren auf landwirtschaftlich relevante Boden- und Klimabedingungen an. Europaweit werden in zwei Stufen 20 bzw. 400 Klassen abgegrenzt, wovon in Deutschland 16 bzw. 184 Klassen vorliegen.

Im **empirischen Vergleich** weisen die Raumgliederungen nach BKR und HYRAUM (Großräume) das größte Potenzial als erklärende Variablen für die Nitratbelastung auf ($R^2_{adj} \geq 0,13$). LBM und AEA können ergänzend relevante Informationen enthalten ($R^2_{adj} \geq 0,06$). Die Geländehöhe ist allenfalls von untergeordneter Bedeutung, kann jedoch in der prozessbezogenen Merkmalsextraktion einen Beitrag leisten (Abschnitt 3.4). Aufgrund der großen Anzahl an BKR-Klassen wurde diese Variable in den in dieser Arbeit durchgeführten Analysen nicht berücksichtigt.

Tabelle 4: Datengrundlagen: 2. Prozessbezogene Umweltvariablen

Variable	Maßstab / Auflösung	Merkmalstyp	Zusammenhang (R^2_{adj})	Nutzen der Interaktion mit HYRAUM
BGR GWN	1 km	Quantitativ	0,010	+0,036
AGRUM GWN	1 km	Quantitativ	0,002	+0,027
AGRUM Abstandsgeschw.	1 km	Quantitativ	0,035	+0,027
AGRUM Verweildauer	1 km	Quantitativ	0,002	+0,033
BÜK1000 nFK	250 m	Quantitativ	0,001	+0,007
HÜK k_f -Wert	1:250.000	Kategorial: 5	0,031	+0,012
HÜK Leitercharakter	1:250.000	Kategorial: 3	0,009	+0,010
HÜK Geochem. Gesteinstyp	1:250.000	Kategorial: 10	0,060	+0,011
BGR Humusgehalt	1:1.000.000	Ordinal: 10	0,006	+0,024

Als Maß für die Stärke des Zusammenhangs dient der korrigierte R^2 (R^2_{adj}). Der Nutzen der Interaktion mit HYRAUM beschreibt die Steigerung des R^2 durch Verwendung des Interaktionsterms. Logarithmisch transformierte Variablen wurden berücksichtigt, wo es angemessen erschien.

3.3 Prozessbezogene Umweltvariablen

Die modellierte GWN, Abstandsgeschwindigkeit und Verweildauer geben wertvolle Auskünfte über die Effektivität des Zuflusses von Nitratemissionen zu einer Messstelle, weshalb sie in hydrogeochemischen Studien häufig als erklärende Variablen genutzt werden. Die genannten Prozessparameter stehen in 1 km Auflösung aus Modellierungen von AGRUM-DE (Zinnbauer et al., 2023) und BGR (2019) zur Verfügung. Insbesondere die Abstandsgeschwindigkeit weist eine relevante Vorhersagegüte auf, nach hydrogeologischen Großräumen differenziert auch die GWN nach BGR und die Verweildauer nach AGRUM-DE (Tabelle 4).

Weitere Prozessparameter sind in der hydrogeologischen Übersichtskarte (HÜK) im Maßstab 1:250.000 (etwa 125 m Auflösung) enthalten, jedoch in kategorialen bzw. partiell ordinalen Variablen. Vor allem der geochemische Gesteinstyp und der k_f -Wert gehen mit Unterschieden in der Nitratkonzentration einher.

Die edaphischen Merkmale nFK und Humusgehalt leisten keinen nennenswerten Beitrag dazu, räumliche Muster in Nitratkonzentrationen zu erklären.

3.4 Prozessbezogene Merkmalsextraktion

Geographische Informationssysteme verfügen über Algorithmen zur Ableitung und Durchführung von Berechnungen innerhalb von topographisch determinierten EZG (vgl. De Smith et al., 2007). Diese erlauben es, aus den oben beschriebenen lokalen Rohdaten prozessbezogene Geoinformationen abzuleiten, die die Verhältnisse im Zustrombereich eines Standorts charakterisieren. Idealerweise sollten hierzu Grundwassergleichenpläne verwendet werden. Da diese zum jetzigen Zeitpunkt nicht bundesweit zur Verfügung stehen, wurde behelfsweise das Copernicus DEM als Proxy herangezogen. Im Gegensatz hierzu werden einfache „Moving-Window-“ Ansätze, bei denen etwa Mittelwerte über omnidirektionale Umgebungen gebildet werden (z.B. bei Knoll et al., 2020), als weniger geeignet angesehen.

Tabelle 5: Datengrundlagen: 3. Prozessbezogene Merkmalsextraktion

Variable	Maßstab / Auflösung	Merkmalstyp	Zusammenhang (R^2_{adj})	Nutzen der Interaktion mit HYRAUM
COP-DEM Topographisches EZG	25 m	Quantitativ	0,009	+0,000
Landwirtschaftlicher Flächenanteil im EZG	25 m	Quantitativ	0,056	+0,079
AGRUM mittlere GWN im EZG	1 km	Quantitativ	0,001	+0,026
AGRUM mittlere Abstandsgeschw. im EZG	1 km	Quantitativ	0,003	+0,009
AGRUM mittlere Verweilzeit im EZG	1 km	Quantitativ	0,001	+0,023

Als Maß für die Stärke des Zusammenhangs dient der korrigierte R^2 (R^2_{adj}). Der Nutzen der Interaktion mit HYRAUM beschreibt die Steigerung des R^2_{adj} durch Verwendung des Interaktionsterms.

Der Nutzen der EZG-bezogenen Herangehensweise wird vor allem beim Attribut „landwirtschaftlicher Flächenanteil im EZG“ deutlich. Dieses leistet nicht nur einen relevanten Beitrag zur räumlichen Trendmodellierung, sondern sein Beitrag übersteigt auch den der (lokalen) Landbedeckungsklasse selbst. Eine Übertragung auf weitere als die in Tabelle 5 gezeigten quantitativen und kategorialen Ausgangsmerkmale erscheint erstrebenswert, vor allem wenn dies auf der Grundlage von Grundwassergleichenplänen anstelle topographischer EZG erfolgen kann.

4 Bewertung aktuell verwendeter Verfahren und ihrer Weiterentwicklungsmöglichkeiten

Ausgehend von der in Abschnitt 2 vorgestellten Problembeschreibung und den daraus abgeleiteten Kriterien und Anforderungen sowie unter Berücksichtigung der zu nutzenden Datengrundlagen aus Abschnitt 3 ist es nun möglich, die bislang verwendeten und/oder in der AVV GeA genannten Regionalisierungsverfahren hinsichtlich ihrer Eignung zu beurteilen (Ziel 1 des Gutachtens). Hierbei sollen auch Weiterentwicklungsmöglichkeiten berücksichtigt werden.

4.1 Voronoi-Verfahren (Nächster-Nachbar-Interpolation, NNI)

Unter „Voronoi-Verfahren“ wird im Zusammenhang mit der AVV GeA allgemein ein Verfahren verstanden, welches sich präziser als Nächster-Nachbar-Interpolation (NNI) bezeichnen ließe, um mögliche Verwechslungen mit der „Voronoi-Interpolation“ oder Natürliche-Nachbarn-Interpolation zu vermeiden. Die AVV GeA verwendet jedoch auch den Begriff „Voronoi-Interpolation“ oder einfach „Voronoi“; es wird hier davon ausgegangen, dass damit ebenfalls das Voronoi-Verfahren, d.h. die NNI gemeint ist.

Bei diesem Verfahren wird einer Zielstelle jener Messwert zugewiesen, der in kürzester (euklidischer) Entfernung zu ihm gemessen wurde. Somit handelt es sich um ein „exaktes“ Interpolationsverfahren, d.h. die interpolierten Werte entsprechen an den Messstellen exakt den beobachteten Werten. Es garantiert grundsätzlich unverzerrte Vorhersagen, sofern die zu regionalisierende Variable keinen systematischen Trend aufweist.

NNI erzeugt ein aus polygonförmigen Stufen bestehendes Interpolationsergebnis (Abbildung 12). Dieses ist im Kontext der Interpolation von Grundwassereigenschaften physikalisch unplausibel, da die „Stufengrenzen“ nicht mit möglichen hydraulischen Barrieren übereinstimmen, sondern sich allein aus der Lage der Messstellen ergeben.

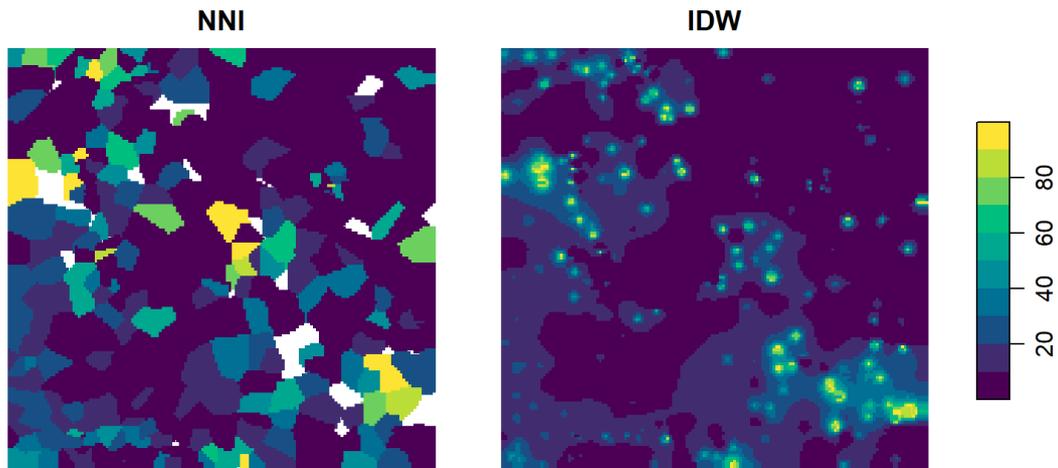
Da NNI ferner innerhalb der Voronoi-Polygone einen konstanten Wert interpoliert, trifft es ferner die implizite Annahme einer ansonsten außerordentlich starken räumlichen Autokorrelation. Diese ist nicht nur empirisch widerlegt (siehe z.B. Abschnitt 2.1), sondern physikalisch ebenso wenig plausibel und im Widerspruch zu den modellierten Unstetigkeiten an Polygongrenzen.

NNI berücksichtigt weiterhin keine erklärenden Variablen und damit keine relevanten Umweltzustandsvariablen.

NNI muss laut AVV GeA bei geringer Messstellendichte von weniger als einer Messstelle je 50 km² angewendet werden. In diesen Situationen mit großer Vorhersagedistanz ist es jedoch gerade nicht sinnvoll, weit entfernte, als unkorreliert anzusehende Messstellenwerte unverändert als Vorhersagen zu übertragen.

Insgesamt erscheint NNI somit auch in datenarmen Situationen wenig geeignet.

Abbildung 12: Regionalisierung von Nitratkonzentrationen in mg/l im Pilotgebiet mit Nächster-Nachbar-Interpolation (NNI, Voronoi-Verfahren) und Inverser Distanzgewichtung (IDW)



Die Regionalisierungen wurden für Box-Cox-transformierte Nitratkonzentrationen durchgeführt und anschließend rücktransformiert.

Quelle: eigene Darstellung, A. Brenning.

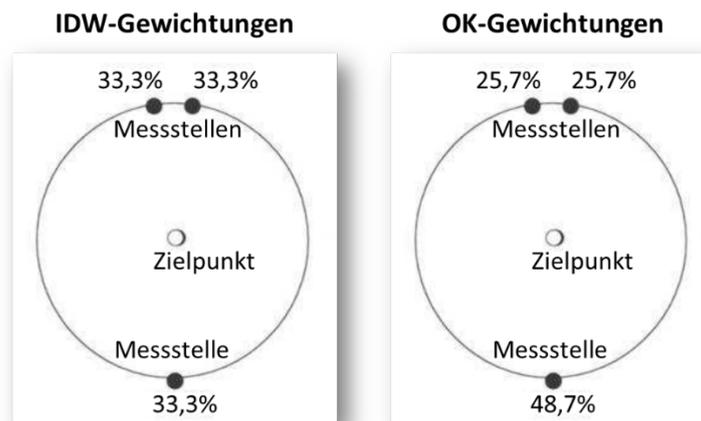
4.2 Inverse Distanzgewichtung (IDW)

Die IDW erzeugt Vorhersagen, indem sie umliegende Beobachtungen proportional zum Kehrwert ihrer Entfernungen gewichtet, wobei keine Zusatzinformationen einbezogen werden. Das Ergebnis ist eine stetige Interpolation, die gerade bei schief verteilten Daten auffällige kreisförmige Artefakte um die Beobachtungsstellen erzeugt (Abbildung 12).

IDW ist ein etabliertes „exaktes“ Interpolationsverfahren. Grundsätzlich erzielt es eine unverzerrte Vorhersage der zu regionalisierenden Variable, sofern diese keinen systematischen Trend aufweist und die Zielvariable nicht transformiert wird. Eine Schwäche der IDW im Vergleich zum OK liegt jedoch darin, dass es die Lagekonfiguration von Messstellen zueinander nicht berücksichtigt – lediglich die euklidische Distanz zum Zielpunkt wird zur Berechnung der Gewichte herangezogen. IDW ist somit im Gegensatz zu Kriging-Verfahren (vgl. Wackernagel, 2003) nicht in der Lage, Abschwächungseffekte zu modellieren. Beispielsweise weist IDW in der in Abbildung 13 skizzierten Vorhersagesituation allen Messstellen das gleiche Gewicht von 33,3% zu, wodurch die beiden nahe beieinander liegenden Messstellen, welche als Pseudoreplikationen angesehen werden können, zusammen mit 66,7% gewichtet werden.

Vor diesem Hintergrund erscheint eine Nutzung von IDW aus methodischer Sicht nicht nachvollziehbar, sobald hinreichend Daten vorliegen, um ein Semivariogramm wenigstens näherungsweise zu schätzen (idealerweise >100 Beobachtungen, vgl. z.B. Webster & Oliver, 2007).

Abbildung 13: Einfluss der Lage von Messstellen bei IDW und OK



Schematisches Beispiel: Während OK die Gewichte der Beobachtungen in etwa hälftig auf die einzelne Beobachtung und das Beobachtungspaar verteilt, weist IDW jeder der Beobachtungen das gleiche Gewicht zu. IDW berücksichtigt somit nicht, dass die beiden nahe beieinander liegenden Beobachtungen fast identische Informationen enthalten. Genaue OK-Gewichte hängen vom konkreten Semivariogramm ab.

Quelle: eigene Darstellung, A. Brenning, verändert und erweitert nach Wackernagel (2003).

4.3 G-FlaG

Die Geostatistische Flächenanteilsschätzung zur Zustandsbewertung von GWK (G-FlaG) ist ein basierend auf der Methode von UBA (2000) durch Beak Consultants (2021) umgesetztes Verfahren, welches zuletzt durch Schafmeister et al. (2023) vorgestellt wurde.

Dieses mit dem Block-Kriging verwandte Verfahren vollzieht den Skalenwechsel (siehe auch Abschnitt 2.6.5) von Punktmessungen auf Voronoi-Polygone innerhalb von GWK. Zielgröße ist der Flächenanteil, der den Schwellenwert überschreitet. Dabei wird kein konkretes Teilgebiet ausgewiesen, sondern der betroffene Flächenanteil innerhalb vorgegebener Teilgebiete geschätzt.

Für die Berechnung der Überschreitungsfläche wird ferner eine Normalverteilungsannahme getroffen. Die Daten werden durch Schafmeister et al. (2023) logarithmisch transformiert. Zwar wird hierdurch augenscheinlich keine Normalverteilung erzielt (Abbildung 5b bei Schafmeister et al., 2023), jedoch verweisen die Autoren auf mögliche alternative Transformationen. Es kann hier nur darüber spekuliert werden, ob das Verteilungsproblem möglicherweise durch die Mittelung auf Blockebene geheilt wird oder zu Verzerrungen führt.

Das G-FlaG-Verfahren bezieht keine erklärenden Variablen mit ein und ist somit nicht in der Lage, relevante Umweltzustandsvariablen und Prozesswissen zu berücksichtigen.

Das verwendete Semivariogramm-Modell wird an die Gesamtheit der Messdaten der Region angepasst. Dies ist eine pragmatische, an sich nachvollziehbare Entscheidung, die jedoch in gewissem Widerspruch zur Nutzung von GWK-Grenzen als „harten“ Barrieren steht (siehe hierzu jedoch Abschnitt 2.6.1).

4.4 SIMIK+

SIMIK+ ist ein spezielles Regressionskriging-Verfahren mit (ausschließlich) kategorialen erklärenden Variablen sowie Varianzheterogenität. Es wurde von Bárdossy et al. (2003) vorgeschlagen und zuletzt als ArcGIS-Toolbox implementiert (Björnsen Beratende Ingenieure, 2022).

SIMIK+ kann als eine pragmatische Spielart des allgemeineren und deutlich besser etablierten KED angesehen werden. Das von Bárdossy et al. (2003) vorgebrachte Argument, KED könne kategoriale Variablen nur über eine willkürliche Numerisierung berücksichtigen, ist aufgrund des bekannten Hilfsmittels der Indikatorvariablen („Dummy-Variablen“) unzutreffend.

SIMIK+ kann sowohl für die Regionalisierung von Messwerten als auch für die Modellierung von Überschreitungswahrscheinlichkeiten per Indikator-Kriging genutzt werden. Derzeit kommt in Thüringen das (numerische) Regionalisierungsverfahren von SIMIK+ zum Einsatz.

SIMIK+ unterstützt – jedenfalls in seiner aktuellen Implementierung – lediglich zwei kategoriale erklärende Variablen, wodurch Umweltzustandsvariablen nur sehr eingeschränkt berücksichtigt werden können. Dabei findet datengetrieben eine Art Variablenselektion bzw. Gruppierung der Kategorien statt. Diese beruht auf einem KV-Fehler.

Innerhalb der auf diese Weise definierten Teilgebiete werden unterschiedliche Varianzen der Messwerte modelliert, wodurch Varianzheterogenität als spezielle Art der Nichtstationarität zugelassen wird (vgl. Abschnitt 2.6.1). Demgegenüber wird das Semivariogramm jedoch "global", also über das gesamte Arbeitsgebiet hinweg homogen modelliert. GWK werden hierbei nicht als Barrieren angesehen.

Die Semivariogramm-Modellierung nach Bárdossy et al. (2003) und der zugehörigen Software-Dokumentation (Usländer, 2003; Bjørnsen Beratende Ingenieure, 2022) zieht die Zielvariable selbst heran. Dies ist bemerkenswert – und mathematisch falsch –, da das zu verwendende Semivariogramm lediglich die nach Berücksichtigung der Teilgebiets-Mittelwerte verbleibende Variabilität charakterisieren müsste. Im Regressionskriging, und dazu gehört auch SIMIK+, ist es daher zwingend erforderlich, nur das Semivariogramm der Residuen des linearen Trendmodells zu modellieren (vgl. z.B. Cressie, 1993).

Über die eigentliche Regionalisierung hinaus bietet die aktuelle SIMIK+-Implementierung ferner als Vorprozessierungsschritt eine automatisierte Ausreißerermittlung und -entfernung an. Diese beruht auf einer Normalverteilungsannahme, welche jedoch im allgemein bei Nitratkonzentrationen unzutreffend ist (vgl. Abschnitt 2.6.2). Die Ausreißerermittlung neigt daher bei der rechtsschiefen Verteilung von Nitratgehalten ausgerechnet zum Abschneiden hoher Nitratgehalte, wodurch tatsächlich nitratbelastete Teilgebiete bei der Regionalisierung unerkannt bleiben könnten. Der SIMIK+-Ausreißertest sollte daher ignoriert werden. (Nach Auskunft von Frau Peters vom Thüringer Landesamt für Umwelt, Bergbau und Naturschutz wird der Ausreißertest daher derzeit in Thüringen lediglich als grober Hinweisgeber für eine manuelle Überprüfung verwendet.)

Unterm Strich lassen sich SIMIK+ durch die grundsätzliche Berücksichtigung von räumlicher Autokorrelation und erklärenden Variablen einige positive Aspekte abgewinnen, die dieses Verfahren deutlich gegenüber NNI und IDW absetzen. Insgesamt ist das Verfahren jedoch zu sehr eingeschränkt, und die identifizierten Mängel sind vermeidbar.

5 Empfehlungen für eine bundesweite Regionalisierung

5.1 Anforderungen an Regionalisierungsverfahren für die Gebietsausweisung

Auf Grundlage der obigen Problemanalysen lässt sich feststellen, dass Regionalisierungsverfahren für die Gebietsausweisung folgende Eigenschaften erfüllen müssen:

1. Regionalisierungsverfahren müssen in der Lage sein, Überschreitungs-Wahrscheinlichkeiten – und nicht lediglich erwartete Nitratgehalte – unter Verwendung geeigneter Wahrscheinlichkeitsmodelle räumlich vorherzusagen. Die Schiefe der Verteilung ist adäquat zu berücksichtigen.
2. Sie müssen hierbei in der Lage sein, erklärende Variablen unterschiedlicher Merkmalstypen einzubeziehen, um räumliche Muster und Trends prozessadäquat und verursachergerecht zu berücksichtigen. Die Möglichkeit von nichtlinearen Beziehungen und Interaktionen zwischen Variablen ist dabei zu beachten.
3. Weitere Heterogenität wurde in Bezug auf die Struktur der räumlichen Abhängigkeiten nachgewiesen (nichtstationäres Semivariogramm); dies ist adäquat zu modellieren.
4. Des Weiteren ist der Skalenwechsel von Punktmessungen zu Kleinflächen (Rasterzellen) zu modellieren, wobei Zielmaßstab oder -auflösung kritisch hinterfragt werden sollten.

Im Rahmen dieser Begutachtung konnte nicht geklärt werden, ab welchem Überschreitungswahrscheinlichkeitsschwellenwert die Ausweisung belasteter Flächen erfolgen sollte. Es wurden Verfahrensweisen eruiert und darauf hingewiesen, dass die naheliegend erscheinende Verwendung eines 50%-Schwellenwertes dazu führt, dass die ausgewiesene Gesamtfläche deutlich kleiner ist als die statistisch abgesicherte Größe der Überschreitungsfläche. Eine weitere Diskussion unter Berücksichtigung von Aspekten des Verwaltungsrechts und des Grundwasserschutzes wird empfohlen.

5.2 Empfehlungen hinsichtlich bislang verwendeter Verfahren

Im Hinblick auf Ziel 1 dieses Gutachtens wird festgestellt, dass von den in der AVV GeA betrachteten Verfahren IDW und NNI („Voronoi“) sowie der Verfahrensgruppe der geostatistischen Methoden lediglich letztere den oben genannten Anforderungen genügen. Dies gilt unabhängig von Schwellenwerten für die Messstellendichte gemäß AVV GeA, welche aus geostatistischer Sicht nicht nachvollziehbar sind und insbesondere aus fachlicher Sicht keineswegs geeignet sind, einen Vorzug für IDW oder NNI gegenüber geostatistischen Verfahren zu rechtfertigen. Eine Verwendung von IDW oder NNI kann aus wissenschaftlicher Sicht nicht empfohlen werden. Auch eine Methodenweiterentwicklung (Ziel 2 der Begutachtung) im Hinblick auf die formulierten Anforderungen ist hier nicht vorstellbar.

Das einzige bisher verwendete geostatistische Verfahren, SIMIK+, weist zahlreiche Einschränkungen und sogar technische Fehler auf. Eine Weiterentwicklung erscheint angesichts der Existenz übergeordneter, gut etablierter und erprobter geostatistischer Verfahren (siehe Abschnitt 2) wenig sinnvoll.

5.3 Empfehlungen für Verfahrensauswahl und -weiterentwicklung

Eine mögliche Vorgehensweise im Bereich der geostatistischen Regionalisierung kann folgende Schritte bzw. Modellkomponenten enthalten, welche allesamt in ihren jeweiligen Anwendungsbereichen etablierte und erprobte Methoden darstellen (Ziel 3 der Begutachtung). Es handelt sich hierbei um einen möglichen und auch kurzfristig deutschlandweit umsetzbaren Ansatz –

alternative Vorgehensweisen beispielsweise mit Ansätzen der Bayes'schen Geostatistik wären denkbar:

1. **Datentransformation** mit Hilfe der Box-Cox-Transformation (vgl. Abschnitt 2.6.2)
2. **Merkmalsextraktion** zur Berücksichtigung hydrogeologischer und geökologischer Verhältnisse im Zustrombereich (Abschnitt 3.4) sowie der Unschärfe der Raumgliederung (Abschnitt 2.6.4)
3. **Erstellung eines Trendmodells** durch automatisierte Auswahl linearer erklärender Variablen mit Hilfe empirischer Kriterien (AIC / BIC; LASSO); ggf. Berücksichtigung nichtlinearer Beziehungen durch additive Modellkomponenten (Spline-Transformation) oder Transformationsfunktionen sowie von Interaktionstermen (vgl. Abschnitt 2.2)
4. **Anpassung eines Semivariogramm-Modells** an die Residuen des Trendmodells; dabei Modellauswahl mit Hilfe empirischer Kriterien (AIC / BIC) auch mit Blick auf unterschiedliche Teilräume (vorzugsweise hydrogeologische Großräume; vgl. Abschnitt 2.2)
5. **Räumliche Vorhersage von Überschreitungswahrscheinlichkeiten** basierend auf Kriging-Varianzen oder bedingter geostatistischer Simulation als zentrales Anliegen (vgl. Abschnitte 2.4 und 2.5)
6. **Schätzung der Gesamtfläche der Überschreitungsgregion** mit Hilfe der bedingten geostatistischen Simulation oder gleichwertiger Verfahren zur Unterstützung der Schwellenwertfindung für eine Gebietsausweisung (vgl. Abschnitt 2.4)
7. **Ggf. Berücksichtigung des Skalenwechsels** von Punktmessungen zu Rasterzellen durch Block-Kriging (bzw. Block-Simulation; vgl. Abschnitt 2.6.5)

Hinsichtlich der Punkte 3 und 4 ist anzumerken, dass eine getrennte Modellierung **hydrogeologischer Großräume** zweckmäßig und praktikabel erscheint. Im Gegensatz hierzu wird empfohlen, von einer separaten Modellierung der Nitratbelastung auf Bundeslandebene abzusehen, da es sich bei dieser administrativen Unterteilung nicht um eine prozessadäquate Untergliederung handelt.

Es wird weiterhin empfohlen, besagte geostatistische Modelle heranzuziehen, um die weitere **Verdichtung des Messstellennetzes** zu optimieren und mit den vorhandenen Ressourcen einen bestmöglichen Nutzen zu erzielen. Auch hier sind nicht Interpolationsunsicherheiten, sondern Unsicherheiten hinsichtlich der Überschreitungswahrscheinlichkeit heranzuziehen, um der Problemstellung der Gebietsausweisung gerecht zu werden.

5.4 Empfehlungen zu Methoden des maschinellen Lernens

Zwar fallen Verfahren des maschinellen Lernens in üblicher Auslegung nicht unter das von der AVV GeA bemühte Konzept der geostatistischen Regionalisierung, jedoch existieren spannende Verfahren an der Schnittstelle zwischen diesen Gebieten, die im weiteren Sinne als geostatistische Verfahren betrachtet werden können. Derzeit sind diese Verfahren gerade mit Blick auf die Modellierung von Überschreitungswahrscheinlichkeiten im Gegensatz zu den etablierten geostatistischen Verfahren noch wenig erforscht. Ihre teils postulierte Überlegenheit ist nicht belastbar dokumentiert, da die verwendeten Validierungsstrategien die Vorhersagesituation oft nicht adäquat nachbilden oder Vergleiche mit geeigneten geostatistischen Modellen in durchgeführt werden. Die hier vorgelegten Untersuchungen liefern keine Hinweise auf relevante Verbesserungen, jedoch sind hier weitere Untersuchungen erforderlich.

Generell stellt sich ferner die Frage, welchen Stellenwert im Rahmen einer Gebietsausweisung die Erwartung haben soll, dass Entscheidungen der öffentlichen Verwaltung transparent und nachvollziehbar sein sollten, und inwiefern derzeit vorhandene Interpretationswerkzeuge für Black-Box-Verfahren des maschinellen Lernens solche Erwartungen erfüllen können (Murdoch

et al., 2019; Rudin, 2019). Trendmodelle, die (mindestens) additiv sind und somit inhärent interpretierbar, entfalten hier ihre Vorzüge (Rudin, 2019).

Vor dem Hintergrund dieser Unsicherheiten kann jedenfalls kurzfristig keine Empfehlung für die Anwendung von Verfahren des maschinellen Lernens für eine operationelle Gebietsausweisung ausgesprochen werden. Sehr wohl empfohlen wird eine tiefergehende Untersuchung und Methodenentwicklung in diesem Bereich, gerade mit Blick auf kombinierte Ansätze sowie hybride Modelle.

6 Fazit

Die Ausweisung nitratbelasteter Gebiete stellt spezielle Anforderungen an Regionalisierungsverfahren. Um systematische Unter- oder Überausweisungen zu vermeiden, muss sich die Verfahrensauswahl und -anwendung an Kriterien orientieren, die sich aus der Problemstellung in Verbindung mit dem rechtlichen Rahmen der AVV GeA verbindlich ableiten.

Das weite Feld der geostatistischen Regressionskriging-Verfahren bietet einen etablierten methodischen Werkzeugkasten, der es erlaubt, prozessadäquat den Anforderungen zu genügen und transparente Entscheidungen zu ermöglichen. Diese Modelle bestehen aus einer linearen, potenziell nichtlinear erweiterbaren räumlichen Trendkomponente sowie einer Kriging-Interpolationskomponente. Heterogenität in hydrogeologischen Verhältnissen kann dabei in beiden Modellkomponenten repräsentiert werden. Dank der mathematisch fundierten Modellstruktur mit überprüfbareren Modellannahmen ist es möglich, den Überschreitungs-Flächenanteil der Nitratbelastung >50 mg/l unverzerrt zu schätzen und dieses Wissen in der Entscheidungsfindung zu nutzen. Diese geostatistischen Modelle weisen auch einen Weg auf, um durch eine an tatsächlichen Ausweisungsunsicherheiten orientierte Messnetzverdichtung den bestmöglichen Nutzen aus kostenträchtigen zusätzlichen Beobachtungen zu ziehen.

Die Nächster-Nachbar-Interpolation („Voronoi-Verfahren“) und die inverse Distanzgewichtung genügen den Anforderungen nicht; sie sollten aus methodischer Sicht auch bei geringer Messstellendichten gemieden werden. Das SIMIK+-Verfahren ist hingegen ein geostatistischer Regressionskriging-Ansatz; aufgrund seiner Einschränkungen kann ebenfalls keine Empfehlung für diese Methode ausgesprochen werden.

Kombinierte und hybride Modelle des maschinellen Lernens mit geostatistischen Komponenten machen derzeit einen bedeutenden Schritt nach vorne. Dennoch liegt neben der mangelnden Modelltransparenz eine große Herausforderung darin, die formulierten Anforderungen zu erfüllen. Nach bisherigem Kenntnisstand stehen zusätzlicher Aufwand, Unsicherheiten und mangelnde Interpretierbarkeit derzeit in keinem guten Verhältnis zum möglichen, aber noch nicht gesicherten Gewinn an Vorhersagegüte, weshalb zunächst weitere Forschung an diesen Verfahren empfohlen wird.

7 Quellenverzeichnis

- AIRBUS (2020): Copernicus DEM: Copernicus digital elevation model product handbook. Report AO/1-9422/18/I-LG, European Space Agency.
- Bárdossy, A.; Giese, H.; Grimm-Strele, J.; Barufke, K.-P. (2003): SIMIK+ — GIS-implementierte Interpolation von Grundwasserparametern mit Hilfe von Landnutzungs- und Geologiedaten. *Hydrologie und Wasserbewirtschaftung*, 47, 13–20. https://doi.org/10.5675/HyWa_2023.1_2.
- Bárdossy, A.; Hörning, S. (2023): Definition of spatial copula based dependence using a family of non-Gaussian spatial random fields. *Water Resources Research*, 59, 7, e2023WR034446. <https://doi.org/10.1029/2023WR034446>.
- Beak Consultants (2021): Identifizierung eines Regionalisierungsverfahrens zur Bewertung des chemischen Zustands von Grundwasserkörpern nach Grundwasserverordnung (GrwV) und Anwendung des Verfahrens in M-V einschließlich Anwendung für die Umsetzung des § 13a Abs. 2 DüV. Unveröffentlichter Abschlussbericht im Auftrag des Landesamtes für Umwelt, Naturschutz und Geologie des Landes Mecklenburg-Vorpommern.
- Beckmann, M.; Didenko, G.; Bullock, J.M.; Cord, A.F.; Paulus, A.; Ziv, G; Václavík, T. (2022): Archetypes of agri-environmental potential: a multi-scale typology for spatial stratification and upscaling in Europe. *Environmental Research Letters*, 17, 11, 115008. <https://doi.org/10.1088/1748-9326/ac9cf5>.
- Beloconi, A.; Chrysoulakis, N.; Lyapustin, A.; Utzinger, J.; Vounatsou, P. (2018): Bayesian geostatistical modeling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. *Environment International*, 121, 1, 57-70. <https://doi.org/10.1016/j.envint.2018.08.041>.
- BGR (2019): Mittlere jährliche Grundwasserneubildung von Deutschland 1:1.000.000 (GWN1000). Digital map data v1. Hannover.
- BGR (2023): Hydrogeologische Raumgliederung von Deutschland. https://www.bgr.bund.de/DE/Themen/Wasser/Projekte/abgeschlossen/Beratung/Hyraum/hyraum_projektbeschr.html (21.06.2023).
- Björnsen Beratende Ingenieure (2022): SIMIK+: ArcGIS-Toolbox zur flächenhaften Darstellung der Grundwasserbeschaffenheit. Unveröffentlichte Programmdokumentation, Thüringer Landesanstalt für Umwelt und Geologie / Björnsen Beratende Ingenieure, Erfurt, 24 S.
- Borgoni, R.; Quatto, P.; Somà, G.; de Bartolo, D. (2010): A geostatistical approach to define guidelines for radon prone area identification. *Statistical Methods and Applications*, 19, 255-276. <https://doi.org/10.1007/s10260-009-0128-x>.
- Boumans, L.; Fraters, D.; van Drecht, G. (2008): Mapping nitrate leaching to upper groundwater in the sandy regions of The Netherlands, using conceptual knowledge. *Environmental Monitoring and Assessment*, 137, 243-249. <https://doi.org/10.1007/s10661-007-9756-5>.
- Burrough, P.A.; McDonnell, R.; Lloyd, C.D. (2015): *Principles of Geographical Information Systems*. 3. Auflage, Oxford University Press, Oxford.
- Breiman, L. (2001): Random forests. *Machine Learning*, 45, 1, 5-32. <https://doi.org/10.1023/A:1010933404324>.
- Brenning, A. (2001): Geostatistics without stationarity assumptions within GIS. *Freiberg Online Geoscience*, Bd. 6, 102 S. <https://doi.org/10.23689/fidgeo-869>.
- Brenning, A. (2012): Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 2012, 5372-5375. <https://doi.org/10.1109/IGARSS.2012.6352393>.

- Brenning, A. (2023): Spatial machine-learning model diagnostics: a model-agnostic distance-based approach. *International Journal of Geographical Information Science*, 37, 3, 584-606. <https://doi.org/10.1080/13658816.2022.2131789>.
- Brenning, A.; Süße, T. (2023): Hinweise zu Trendtests gemäß Grundwasserverordnung. Unveröffentlichte schriftliche Mitteilung an das Umweltbundesamt, 3 S.
- Cressie, N.A.C. (1993): *Statistics for Spatial Data*. Wiley, New York.
- Cressie, N.; Suesse, T. (2020): Great expectations and even greater exceedances from spatially referenced data. *Spatial Statistics*, 37, 100420. <http://dx.doi.org/10.1016/j.spasta.2020.100420>.
- De Smith, M.J.; Goodchild, M.F.; Longley, P.A. (2007): *Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools*. Troubadour, Leicester.
- Erisman, J.W.; Galloway, J.N.; Seitzinger, S.; Bleeker, A.; Dise, N.B.; Petrescu, A.M.R.; Leach, A.M.; de Vries, W. (2013): Consequences of human modification of the global nitrogen cycle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368, 1621, 20130116. <https://doi.org/10.1098/rstb.2013.0116>.
- Fotheringham, A.S.; Brunson, C.; Charlton, M.E. (2002): *Geographically Weighted Regression*. Wiley, Chichester.
- Gnann, S.J.; Allmendinger, M.C.; Haslauer, C.P.; Bárdossy, A. (2018): Improving copula-based spatial interpolation with secondary data. *Spatial Statistics*, 28, 105-127. <https://doi.org/10.1016/j.spasta.2018.07.001>.
- Hawker, L.; Neal, J. (2021): FABDEM V1-0. <https://doi.org/10.5523/bris.25wfy0f9ukoge2gs7a5mqpg2j7>.
- Hawker, L.; Uhe, P.; Paulo, L.; Sosa, J.; Savage, J.; Sampson, C.; Neal, J. (2022): A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17, 024016. <https://doi.org/10.1088/1748-9326/ac4d4f>.
- Helsel, D.R. (2005): *Nondetects and Data Analysis: Statistics for Censored Environmental Data*. Wiley, New York.
- Hooker, G.; Mentch, L.; Zhou, S. (2021): Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31, 82. <https://doi.org/10.1007/s11222-021-10057-z>.
- James, G.; Witten, D.; Hastie, T.; Tibshirani, R. (2021): *An Introduction to Statistical Learning with Applications in R*. Springer, New York.
- Karimanzira, D.; Weis, J.; Wunsch, A.; Ritzau, L.; Liesch, T.; Ohmer, M. (2023): Application of machine learning and deep neural networks for spatial prediction of groundwater nitrate concentration to improve land use management practices. *Frontiers in Water*, 5, 1193142. <https://doi.org/10.3389/frwa.2023.1193142>.
- Knoll, L.; Breuer, L.; Bach, M. (2020): Nation-wide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environmental Research Letters*, 15, 064004. <https://doi.org/10.1088/1748-9326/ab7d5c>.
- Koch, J.; Stisen, S.; Refsgaard, J.C.; Ernstsén, V.; Jakobsen, P.R.; Højberg, A.L. (2019): Modeling depth of the redox interface at high resolution at national scale using random forest and residual Gaussian simulation. *Water Resources Research*, 55, 2, 1451-1469. <https://doi.org/10.1029/2018WR023939>.
- Milà, C.; Mateu, J.; Pebesma, E.; Meyer, H. (2022): Nearest neighbour distance matching leave-one-out cross-validation for map validation. *Methods in Ecology and Evolution*, 13, 6, 1304-1316. <https://doi.org/10.1111/2041-210X.13851>.

- Molnar, C. (2022): Interpretable machine learning — a guide for making black box models explainable. <https://christophm.github.io/interpretable-ml-book/>.
- Murdoch, W.J.; Singh, C.; Kumbier, K.; Yu, B. (2019): Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116, 44, 22071-22080. <https://doi.org/10.1073/pnas.1900654116>.
- Ohlert, P.L.; Bach, M.; Breuer, L. (2023a): Verfahren zur Regionalisierung der Nitratkonzentration im Grundwasser. *Hydrologie und Wasserbewirtschaftung*, 67, 1, 20-33. https://doi.org/10.5675/HyWa_2023.1_2.
- Ohlert, P.L.; Bach, M.; Breuer, L. (2023b): Accuracy assessment of inverse distance weighting interpolation of groundwater nitrate concentrations in Bavaria (Germany). *Environmental Science and Pollution Research*, 30, 9445-9455. <https://doi.org/10.1007/s11356-022-22670-0>.
- Petermann, E.; Meyer, H.; Nussbaum, M.; Bossew, P. (2021): Mapping the geogenic radon potential for Germany using machine learning. *Science of the Total Environment*, 754, 142291. <https://doi.org/10.1016/j.scitotenv.2020.142291>.
- Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat (2019): Deep learning and process understanding for data-driven Earth system science. *Nature*, 566, 195. <https://doi.org/10.1038/s41586-019-0912-1>.
- Roßberg, D.; Michel, V.; Graf, R.; Neukampf, R. (2007): Definition von Boden-Klima-Räumen für die Bundesrepublik Deutschland. *OpenAgrar*, 59, 7, 155–161.
- Rudin, C. (2019): Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215. <https://doi.org/10.1038/s42256-019-0048-x>.
- Schafmeister, M.T.; Steffen, M.; Zeissler, K.O.; Zingelmann, M. (2023): Extension variance: an early geostatistical concept applied to assess nitrate pollution in groundwater. *Hydrogeology Journal*, 31, 1463-1473. <https://doi.org/10.1007/s10040-023-02666-x>.
- Schratz, P.; Becker, M.; Lang, M.; Brenning, A. (2021): mlr3spatiotempcv: spatiotemporal resampling methods for mlr3. *arXiv preprint*, arXiv:2110.12674. <https://doi.org/10.48550/arXiv.2110.12674>.
- Sekulic, A.; Kilibarda, M.; Heuvelink, G.B.M.; Nikolic, M.; Bajat, B. (2020): Random forest spatial interpolation. *Remote Sensing*, 12, 10, 1687. <https://doi.org/10.3390/rs12101687>.
- Tobler, W. (1987): Measuring spatial resolution. In: *Proceedings, Land Resources Information Systems Conference*. Beijing, S. 12-16.
- UBA (2000): Flächenhafte Darstellung punktbezogener Daten über Stoffgehalte in Böden: Workshop, veranstaltet vom Umweltbundesamt, im Auftrag des ständigen Ausschusses Informationsgrundlagen der Bund-Länder-Arbeitsgemeinschaft Boden (LABO). *Texte*, 49/2000, 140-145. Umweltbundesamt, Berlin.
- UBA (2023): Leistungsbeschreibung. Az. 20 032-17/0040, 14.03.2023, 11 S.
- Usländer, T. (2003): Benutzerhandbuch SIMIK+: ArcView-Erweiterung zur flächenhaften Darstellung der Grundwasserbeschaffenheit. Unveröffentlichte Programmdokumentation, Umweltinformationssystem Baden-Württemberg / Fraunhofer-Institut für Informations- und Datenverarbeitung, Karlsruhe, 61 S.
- Webster, R.; Oliver, M.A. (2007): *Geostatistics for Environmental Scientists*. 2. Auflage, Wiley, New York.
- Wackernagel, H. (2003): *Multivariate Geostatistics: An Introduction with Applications*. Springer-Verlag, Berlin. <http://dx.doi.org/10.1007/978-3-662-05294-5>.
- Wriedt, G.; de Vries, D.; Eder, T.; Federolf, C. (2019): Regionalisierte Darstellung der Nitratbelastung im Grundwasser Niedersachsens. *Grundwasser*, 24, 27-41. <https://doi.org/10.1007/s00767-019-00415-0>.

Wood, S.N. (2017): Generalized Additive Models: An Introduction with R. 2. Auflage, Chapman and Hall/CRC Press.

Zinnbauer, M.; Eysholdt, M.; Henseler, M.; Herrmann, F.; Kreins, P.; Kunkel, R.; Nguyen, H.; Tetzlaff, B.; Venohr, M.; Wolters, T.; Wendland, F. (2023): Quantifizierung aktueller und zukünftiger Nährstoffeinträge und Handlungsbedarfe für ein deutschlandweites Nährstoffmanagement – AGRUM-DE Thünen Report, 108, Johann Heinrich von Thünen-Institut, Braunschweig. <https://doi.org/10.3220/REP1684153697000>.