

TEXTE

88/2026

Abschlussbericht

Grundwasservulnerabilitäts- bewertung im Zulassungsverfahren von Pflanzenschutzmitteln (FARM)

von:

Dr. Anne-Karin Cooke, Dr. Kristof Dorau, Ephraim Erkens, Felix Joger, Dr. Sandra Willkommen,
Dr. Stefan Broda

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover, Dienstbereich Berlin

Herausgeber:

Umweltbundesamt

TEXTE 88/2026

REFOPLAN des Bundesministeriums Umwelt,
Naturschutz, nukleare Sicherheit und Verbraucherschutz

Forschungskennzahl 3720 23 401 0

Abschlussbericht

Grundwasservulnerabilitätsbewertung im Zulassungsverfahren von Pflanzenschutzmitteln (FARM)

von

Dr. Anne-Karin Cooke, Dr. Kristof Dorau, Ephraim Erkens,
Felix Joger, Dr. Sandra Willkommen, Dr. Stefan Broda
Bundesanstalt für Geowissenschaften und Rohstoffe (BGR),
Hannover, Dienstbereich Berlin

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:

Bundesanstalt für Geowissenschaften und Rohstoffe (BGR)
Stilleweg 2
30655 Hannover

Dienstbereich Berlin
Wilhelmstraße 26
13593 Berlin

Abschlussdatum:

September 2025

Redaktion:

Fachgebiet Umweltexposition und Grundwasserrisiken Pflanzenschutzmittel (IV 1.3-2)
Dr. Wolfram König, Stephan Marahrens, Mark Rauch

DOI:

<https://doi.org/10.60810/openumwelt-8222>

ISSN 1862-4804

Dessau-Roßlau, Juni 2026

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

Zitiervorschlag:

Cooke, A.-K., Dorau, K., Erkens, E., Joger, F., Willkommen, S., Broda, S. (2026),
Grundwasservulnerabilitätsbewertung im Zulassungsverfahren von Pflanzenschutzmitteln (FARM), In:
Schriftenreihe Umweltbundesamt TEXTE 88/2026, <https://doi.org/10.60810/openumwelt-8222>

Kurzbeschreibung: Grundwasservulnerabilitätsbewertung im ZulassungsverFAhRen von PflanzenschutzMitteln (FARM)

Die harmonisierte Bewertung der Vulnerabilität des Grundwassers in Bezug auf Einträge von Wirkstoffen und ihren Abbauprodukten wird im Zulassungsverfahren von Pflanzenschutzmitteln (PSM) relevant, wenn zielgerichtete Monitoringdaten für die Bewertung des Grundwasserrisikos herangezogen werden. In dem Projekt wurden Faktoren der Vulnerabilität anhand von Daten aus den Überwachungsprogrammen der Bundesländer untersucht. Informationen und Daten zu Wirkstoffen und Metaboliten von PSM im Grundwasser wurden in einer umfangreichen bundesweiten Datenbank für Deutschland zusammengeführt, harmonisiert und ausgewertet. Zur räumlichen Vorhersage von Stoffkonzentrationen an Messstellen und um Standortfaktoren als relevante Treiber der Vulnerabilität von Grundwasserleitern gegenüber dem PSM-Eintrag zu identifizieren, wurden Modelle des maschinellen Lernens (ML) anhand der genannten Datengrundlage trainiert. Die ML-Modelle waren in der Lage, relevante Faktoren und räumliche Zusammenhänge aufzudecken, insbesondere der Einfluss der Landwirtschaft und der Filtertiefe wurden als steuernde Größen identifiziert. Für acht mobile und sehr mobile Abbauprodukte von drei Wirkstoffen stehen oberflächennahe Filter und größere Anbauflächen der zugehörigen Kulturen im Umkreis der Messstellen im Zusammenhang mit höheren Konzentrationen im Grundwasser. Weitere erklärende Variablen (Features) wie Bodeneigenschaften (Sandgehalt, Makroporen) und die Grundwasserneubildung zeigten je nach Stoff z.T. unterschiedliche Effekte auf die Stoffkonzentrationen. Hydrogeologische Variablen spielten in den Metabolitenmodellen eine untergeordnete Rolle. Es ist davon auszugehen, dass die Modelle vorwiegend lernen, welche Gebiete wahrscheinlich einen PSM-Einsatz aufweisen. Dies basiert zum einen auf räumlichen Korrelationen, die der Verteilung der kulturspezifischen Anbauflächen innerhalb Deutschlands geschuldet sind. Die Bodeneigenschaften, die Höhenlage, der Niederschlag und die Nitratkonzentrationen sind solche Parameter, die eher auf eine Korrelation mit Gebieten hinweisen, in denen eine bestimmte Kultur verstärkt angebaut wird, sowie generell auf landwirtschaftliche Nutzung. Aufgrund fehlender tatsächlicher Anwendungsdaten von Wirkstoffen auf landwirtschaftlichen Flächen wird die Ableitung kausaler Zusammenhänge in Bezug auf die Grundwasservulnerabilität erschwert, weshalb die Modelle nicht für eine Indexentwicklung genutzt werden konnten. Die Ergebnisse beziehen sich auf mobile Metaboliten mit langer Anwendungshistorie ihrer Muttersubstanzen und damit auf ein begrenztes Stoffspektrum. Wirkstoffmodelle und Wirkstoffgruppenmodelle konnten aufgrund des hohen Anteils der Messungen unterhalb der Bestimmungsgrenze nur in einem begrenzten Rahmen getestet werden. Modelle von sehr mobilen und mobilen Wirkstoffen zeigten eine den Metabolitenmodellen ähnliche Gewichtung der Variablen. Aus einem Abgleich zwischen den Monitoringdaten ausgewählter Stoffe und etablierten Vulnerabilitätsverfahren (Schutzfunktion der Grundwasserüberdeckung nach Hölting et al., 1995 und LAWA; sowie zwei Konfigurationen des Prozessmodells GeoPELMO DE) lässt sich keine klare Empfehlung für eine der beiden Methoden ableiten. Aus diesem Grund sollte sich die Auswahl von Standorten für zielgerichtete Nachzulassungsmonitoringstudien zum gegenwärtigen Zeitpunkt ausdrücklich auf das Urteil von Experten stützen, wobei lokale Daten mit höherer räumlicher Auflösung verfügbar sein sollten.

Abstract: Groundwater vulnerability assessment in the context of plant protection product authorizations (FARM)

The harmonized assessment of groundwater vulnerability according to entries of active substances and their degradation products becomes relevant for the authorization of plant protection products (PPPs) if targeted monitoring data is used in the groundwater risk assessment. In the project, vulnerability factors were investigated using data from the monitoring programs of the federal states. Information and data of active substances and metabolites of PPP in groundwater was assembled, harmonized, and analyzed in a nationwide database for Germany. Machine learning (ML) models were trained with a variety of explanatory features with the aim to predict PPP concentration classes at monitoring sites and to assess the drivers that characterize the vulnerability of the aquifers. The ML models were capable to learn relevant factors and spatial relationships, especially agricultural intensity and the top filter edge were identified as key features. Eight mobile and very mobile transformation products from three active substances were strongly affected by shallow filter depths and the size of the cultivated area of the main crops in vicinity of the monitoring well showing increased PPP concentrations in groundwater. Further explanatory variables (features), such as soil properties (sand content and macropore classes) and groundwater recharge, demonstrated compound specific partly different effects on PPP concentrations in groundwater. Hydrogeological features played only a minor role for the metabolite models. We assume that the ML models predominantly learned which areas are likely to be exposed to PPP use. This is based, on the one hand, on spatial dependencies driven by the distribution of the crop-specific area under cultivation. Soil characteristics, altitude, precipitation, and nitrate concentrations are such parameters which are an indicator for regions with a distinct crop and with agricultural use intensity in general. Because of missing actual application data causal relations in relation to groundwater vulnerability are uncertain and building an index upon the empirical data is not possible. *Sensu stricto*, the results are valid for a limited spectrum of compounds (mobile metabolites with long-lasting application history of their parent compounds). Active substance models and active substance group models could barely be tested because of left-censored data below the limit of quantification. Model results of very mobile and mobile active substances and metabolites demonstrated a similar behavior with respect to the feature dependence. A comparison between monitoring data of selected compounds and established vulnerability concepts (protective function of the aquifer according to Hölting et al. (1995) and LAWA; two process-based model versions according to GeoPELMO DE) did not enable to clearly recommend one of the two methods. Because of this, targeted post-registration monitoring studies should explicitly rely on expert judgement for site selection and characteristics with local data availability at higher spatial resolution.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	10
Tabellenverzeichnis.....	21
Abkürzungsverzeichnis.....	22
Zusammenfassung.....	24
Summary.....	30
1 Hintergrund und Zielsetzung.....	36
2 Methoden.....	41
2.1 Datengrundlage und Datenaufbereitung.....	41
2.1.1 Wirkstoff- und Metabolitenkonzentrationen sowie Stammdaten der Grundwassermessstellen.....	41
2.1.2 PSM-Stoffeigenschaften.....	42
2.1.3 Grundwasserchemische Messungen.....	42
2.1.4 PSM-Zulassung und Verkaufszahlen.....	42
2.1.5 Standortfaktoren.....	43
2.1.5.1 Hydrogeologie.....	43
2.1.5.2 Wasserhaushalt.....	43
2.1.5.3 Grundwasserflurabstand.....	43
2.1.5.4 Boden.....	44
2.1.5.5 Klima.....	44
2.1.5.6 Topographie.....	44
2.1.5.7 Gewässer.....	44
2.1.5.8 Landnutzung.....	44
2.1.6 Weitere, nicht verwendete Daten.....	46
2.1.6.1 Geschätzte Aufwandmengen.....	46
2.1.6.2 Maximale Aufwandmengen aus Zulassungssituation und InVeKoS-Flächen.....	46
2.1.6.3 Geschätzte PSM-Aufwandmengen aus PAPA-Erhebungen.....	47
2.2 Verwendete Methoden für die statistische Auswertung.....	48
2.2.1 Methoden für linkszensierte Daten.....	48
2.2.2 Explorative Datenanalyse.....	50
2.2.3 Methoden des maschinellen Lernens.....	51
2.2.3.1 Was ist maschinelles Lernen?.....	51
2.2.3.2 Räumliches vs. Zeitliches Modell.....	51
2.2.3.3 Modellgüte.....	52

2.2.3.4	Im Projekt verwendete Algorithmen	53
2.2.3.5	Feature-Auswahl und Modellvarianten	60
2.2.3.6	Stoffauswahl für das Maschinelle Lernen und Konzentrationsklassen	63
2.2.3.7	Erklärbarkeit und Interpretierbarkeit von Modellen	65
3	Ergebnisse	67
3.1	Auswertung der Datenbank	67
3.1.1	Messprogramme	67
3.1.2	Stoffspektrum	72
3.1.3	Fundsituation	80
3.1.3.1	Gesamtsituation.....	80
3.1.3.2	Stoffliche Auswertung.....	85
3.1.3.3	Zeitlicher Verlauf.....	88
3.1.3.4	Überschreitungsfrequenzen und Verkaufszahlen.....	92
3.1.3.5	Der Einfluss von Standortfaktoren auf die Fundsituation.....	92
3.2	Modelle	94
3.2.1	Reduzierte Feature-Auswahl.....	95
3.2.1.1	Vorhersagegüte.....	95
3.2.1.2	Featuregewichtung (Feature Importance).....	98
3.2.1.3	Feature Effekte	103
3.2.1.4	Einfluss der Standortfaktoren in Anbaugebieten.....	113
3.2.1.5	Räumliche Vorhersagen	120
3.2.1.6	Analyse der Residuen.....	128
3.2.1.7	Feature-Interaktionen.....	131
3.2.1.8	Vergleich der Vorhersage zwischen Random-Forest und dem Neuronalen Netz.....	132
3.2.2	Erweiterte Feature-Auswahl	133
3.2.2.1	Vorhersagegüte.....	134
3.2.2.2	Featuregewichtung (Feature Importance).....	136
3.2.2.3	Feature-Effekte	138
3.2.3	Zusammenfassung	142
3.2.4	Wirkstoffmodelle für Stoffgruppen und Einzelstoffe.....	142
3.2.4.1	Reduzierte Feature-Auswahl.....	143
3.2.4.2	Erweiterte Feature-Auswahl	148
3.2.4.3	Zusammenfassung	151
4	Diskussion.....	152

4.1	Metaboliten-Modelle	152
4.1.1	Reduzierte Feature-Auswahl.....	152
4.1.1.1	Konzentrationsklassen und Vorhersagegüte	152
4.1.1.2	Feature-Gewichtung und Feature-Effekte der reduzierten Feature-Auswahl	153
4.1.2	Erweiterte Feature-Auswahl	158
4.2	Wirkstoff- und Wirkstoffgruppenmodelle.....	159
4.3	Diskussion zur Datenverfügbarkeit	160
5	Bewertung der Vulnerabilität.....	162
5.1	Verfahren zur Bestimmung der intrinsischen Vulnerabilität.....	162
5.1.1	Hintergrund.....	162
5.1.2	Abgleich mit Monitoringdaten.....	163
5.2	GEOPELMO DE – Abgleich mit Daten und Vorhersagen	167
5.2.1	Hintergrund.....	167
5.2.2	Abgleich der Simulationen mit Monitoringdaten	168
5.3	Diskussion und Schlussfolgerungen zur Anwendung und Weiterentwicklung von Grundwasser-Vulnerabilitätsindizes	178
6	Nachzulassungsmonitoring	186
6.1	Zielgerichtete Monitoringstudien im Nachzulassungsverfahren (NZM)	186
6.2	Messstellenauswahl nach Kriterien und Datenverfügbarkeit.....	187
6.3	Empfehlungen für zielgerichtete Monitoringstudien.....	193
7	Ausblick	197
8	Quellenverzeichnis	199
A	Anhang	208
A.1	Disclaimer zur Verwendung der Daten des FARM-Projektes und Informationen zur Datennachnutzung.....	208
A.2	Informationen zu den Eingangsdaten (features) die für das maschinelle Lernen verwendet wurden.....	210
A.3	Exkurs zum Einfluss des Klimawandels auf die Grundwasservulnerabilität durch PSM	213
A.4	Zusatzinformationen zu InVeKoS-Daten.	214
A.5	Rechenbeispiele zur Bewertung der Modellgüte.....	216
A.6	Exkurs zu Pilotgebieten	218
A.7	Feature-Effekte der reduzierten Feature-Auswahl in den Neuronalen Netze	219
A.8	Zusammenstellung von Kriterien die für eine Vielzahl von Verfahren für die Vulnerabilitätsbewertung von Grundwasser herangezogen werden.	224

Abbildungsverzeichnis

Abbildung 1:	Beispiel einer Verteilung von Konzentrationen mit linkszentriertem Anteil unterhalb der Nachweis- (LOD) oder Bestimmungsgrenze (LOQ).....	48
Abbildung 2:	Vereinfachtes Schema zur Einteilung der Daten beim Maschinellen Lernen.	51
Abbildung 3:	Schematische Darstellung der Modellarchitektur, die für die Random Forest (RF)-Modelle verwendet wurde. Das Modell-Tuning erfolgte anhand von Bagging unter der Verwendung des OOB als Validierungsset (innere Umrandungen). Die Evaluierung der Modelle mit optimaler Hyperparameter-Kombination und Entscheidungsschwellenwert erfolgte anhand einer 5-fachen Kreuzvalidierung jeweils auf einem unabhängigen Testset (a). Anschließend wurde das Modell-Tuning für den gesamten Datensatz wiederholt (b) und das Modell mit der besten Hyperparameter-Kombination und dem besten Entscheidungsschwellenwert auf den gesamten Daten trainiert (c).	56
Abbildung 4:	Abwägung zwischen der Anzahl an Messstellen und Anzahl an Features mit dem Ziel einer möglichst hohen Modellgüte.	61
Abbildung 5:	Gesamtheit der Messstellen im Datensatz, an denen jemals Messungen von Wirkstoff- bzw. Metabolitenkonzentrationen im Grundwasser durchgeführt wurden.....	68
Abbildung 6:	Anzahl der Messstellen, die in der FARM Datenbank berücksichtigt werden differenziert in Landesmessnetz und Wasserversorgern bzw. Dritten.....	69
Abbildung 7:	Anzahl der Messstellen pro Bundesland über die Zeit.	69
Abbildung 8:	Das durchschnittliche Messintervall in Jahren ist gegen den prozentualen Anteil der Messungen in Form einer kumulativen Verteilungskurve aufgetragen und für den Gesamtzeitraum (1973-2019; schwarze durchgezogene Linie) bzw. für das letzte Jahrzehnt (2010-2019; graue gestrichelte Linie) differenziert dargestellt.	70
Abbildung 9:	Anzahl der Messstellen und Anzahl der insgesamt durchgeführten Messungen; jeweils dargestellt für den Gesamtzeitraum (1973-2019; oben) und für den Zeitraum von 2010-2019 (unten).	70
Abbildung 10:	Anzahl der Messstellen pro Anzahl der Messungen für den Zeitraum 2010-2019 (a) und durchschnittliche Anzahl an Messungen pro Messstellen über Anzahl der Messungen nach Bundesland gruppiert (b) (exemplarisch für Methylphenyl-Chloridazon).	71

Abbildung 11:	Zeitlicher Verlauf der Anzahl der Stoffe im Monitoring Programm pro Jahr der ersten Messung.....	71
Abbildung 12:	Anzahl der im Monitoring untersuchten Wirkstoffe und Metaboliten pro Jahr.....	72
Abbildung 13:	Anzahl der gesuchten Stoffe pro Jahr und Bundesland ab 1973 bis 1999.	73
Abbildung 14:	Anzahl der gesuchten Stoffe pro Jahr und pro Bundesland ab 2000. In den Jahren nach 2010 scheint sich das Messprogramm in den Bundesländern weitestgehend etabliert zu haben.	74
Abbildung 15:	Anzahl der im Monitoring beprobten Stoffe aus dem Gesamtdatensatz pro Bundesland.	75
Abbildung 16:	Barcode-Darstellung des Stoffspektrums pro Bundesland.	75
Abbildung 17:	Anzahl der Stoffe im Monitoring pro Bundesland und Pestizidklasse, für die Messungen > LOQ vorliegen.	77
Abbildung 18:	Anzahl der Stoffe im Datensatz gruppiert nach Mobilitäts- und Persistenzklasse, gesondert dargestellt für Wirkstoffe und relevante Metaboliten bzw. für nicht-relevante Metaboliten.	78
Abbildung 19:	Anzahl der Stoffe nach Zensurgrad (Datensatz ab einschließlich 2010).	79
Abbildung 20:	Anzahl der Stoffe je Anteil der Messstellen mit ausschließlich zensierten Werten.....	79
Abbildung 21:	Anzahl der Stoffe je Anteil der Messstellen, an denen alle Messungen oberhalb der Bestimmungsgrenze liegen.	80
Abbildung 22:	Darstellung der Anzahl von Messstellen mit mindestens einer Überschreitung von $0,1 \mu\text{g L}^{-1}$ im Gesamtdatensatz, jeweils für Wirkstoffe und Metaboliten und aufgeteilt Mobilität und Persistenz.	82
Abbildung 23:	Anzahl der Messstellen pro Bundesland mit mindestens einer Überschreitung von $0,1 \mu\text{g L}^{-1}$ bei Wirkstoffen und relevanten Metaboliten im Zeitraum ab 2010.	83
Abbildung 24:	Räumliche Verteilung der Messstellen, an denen Wirkstoffe und relevanten Metaboliten $0,1 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.	83
Abbildung 25:	Räumliche Verteilung der Messstellen, an denen Wirkstoffe $0,1$ $\mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.....	84
Abbildung 26:	Räumliche Verteilung der Messstellen, an denen nicht- relevanten Metaboliten $3 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.	84
Abbildung 27:	Räumliche Verteilung der Messstellen, an denen nicht- relevanten Metaboliten $10 \mu\text{g L}^{-1}$ mindestens einmal	

	überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.85
Abbildung 28:	Verfügbarkeit von InVeKoS-Daten nach Bundesland und Jahr aufgeschlüsselt. Die Zahlen geben den Anteil an Messstellen mit verfügbaren InVeKoS-Daten in Prozent an.89
Abbildung 29:	Verfügbarkeit der Features in Abhängigkeit der Zeit für Dimethachlor-CA. Die Matrizen zeigen den Anteil fehlender Werte je Feature und Jahr. Die linke Matrix enthält die statischen Features; die rechte Matrix enthält die zeitlich variablen Features. Das Histogramm rechts zeigt die Summe aller Messungen der Stoffkonzentrationen pro Jahr als ungefüllte Balken und die Summe der Messungen an Messstellen, für die alle Features verfügbar sind als ausgefüllte Balken.90
Abbildung 30:	Variation der Konzentrationsklassen bei Desphenyl-Chloridazon in Abhängigkeit der Länge der durchgehenden Zeitreihen.91
Abbildung 31:	Spearman Korrelationskoeffizient rho für die Korrelationen zwischen Standortfaktoren und der Überschreitungsfrequenz von $0,1 \mu\text{g L}^{-1}$93
Abbildung 32:	Die Modellgüte der binären Random-Forest-Modelle der reduzierten Feature-Auswahl für die acht Metaboliten.96
Abbildung 33:	In Ergänzung zu Abbildung 32 wird hier die Modellgüte des Random-Forest Modells für Trifluoressigsäure mit der Klassengrenze $0,5 \mu\text{g L}^{-1}$ gezeigt.97
Abbildung 34:	Gewichtung der Features nach der Permutation Feature Importance für das binäre Random-Forest Modell für die reduzierte Feature Auswahl.99
Abbildung 35:	Gewichtung der Features nach der Gini impurity Feature Importance für das binäre Random Forest Modell für die reduzierte Feature Auswahl.101
Abbildung 36:	Gewichtung der Features nach der Permutation Feature Importance für das binäre Neuronale Netz Modell für die reduzierte Feature Auswahl. Hierbei wird die durchschnittliche Importance, die aus den 5 Teildatensätzen ergibt, gezeigt.102
Abbildung 37:	Gewichtung der Features nach der Permutation Feature Importance für das binäre Neuronale Netz Modell für die reduzierte Feature Auswahl. Hierbei wird die durchschnittliche Importance, die aus den 5 Teildatensätzen ergibt, gezeigt.103
Abbildung 38:	Feature-Effekte von Desphenyl-Chloridazon im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide

	Konzentrationsklassen in der Reihenfolge der Feature-Importance erzeugt.	105
Abbildung 39:	Feature-Effekte von Methyldesphenyl-Chloridazon im binären Random-Forest Modell der reduzierten Feature-Auswahl. ...	106
Abbildung 40:	Feature-Effekte von Metazachlor ESA im binären Random-Forest Modell der reduzierten Feature-Auswahl.	107
Abbildung 41:	Feature-Effekte der Metazachlorsäure im binären Random-Forest Modell der reduzierten Feature-Auswahl.	108
Abbildung 42:	Feature-Effekte von Metolachlor ESA im binären Random-Forest Modell der reduzierten Feature-Auswahl.	109
Abbildung 43:	Feature-Effekte von Metolachlor OXA im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.	110
Abbildung 44:	Feature-Effekte von Metolachlor NOA 413173 im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.	111
Abbildung 45:	Feature-Effekte der Trifluoressigsäure im binären Random-Forest Modell der reduzierten Feature-Auswahl.	112
Abbildung 46:	Verteilung der Werte der Filteroberkante in den Anbaugebieten.	114
Abbildung 47:	Verteilung der Werte des Sandgehalts in den Anbaugebieten.	115
Abbildung 48:	Verteilung der Werte der Grundwasserneubildung in den Anbaugebieten.	116
Abbildung 49:	Verteilung der Kategorien der Hohlraumart in den Anbaugebieten.	117
Abbildung 50:	Verteilung der Werte der Kategorien der Makroporen in den Anbaugebieten.	118
Abbildung 51:	Verteilung der Werte des organischen Kohlenstoffs im Boden in den Anbaugebieten.	119
Abbildung 52:	Verteilung der Werte der Kategorien der Durchlässigkeit des Grundwasserleiters in den Anbaugebieten.	120
Abbildung 53:	Tatsächliche und vorhergesagte Konzentrationsklasse für Desphenyl-Chloridazon aller Grundwassermessstellen.	121
Abbildung 54:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Desphenyl-Chloridazon der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).	121
Abbildung 55:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Methyldesphenyl-Chloridazon der	

	Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	122
Abbildung 56:	Tatsächliche und vorhergesagte Konzentrationsklasse für Metazachlor ESA aller Grundwassermessstellen.	122
Abbildung 57:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metazachlor ESA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	123
Abbildung 58:	Tatsächliche und vorhergesagte Konzentrationsklasse für Metazachlorsäure aller Grundwassermessstellen.	123
Abbildung 59:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metazachlorsäure der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	124
Abbildung 60:	Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor ESA aller Grundwassermessstellen.....	124
Abbildung 61:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor ESA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	125
Abbildung 62:	Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor OXA aller Grundwassermessstellen.....	125
Abbildung 63:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor OXA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	126
Abbildung 64:	Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor-Metabolit NOA 413173 aller Grundwassermessstellen.	126
Abbildung 65:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor-Metabolit NOA 413173 der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).	127
Abbildung 66:	Tatsächliche und vorhergesagte Konzentrationsklasse für Trifluoressigsäure aller Grundwassermessstellen.....	127
Abbildung 67:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Trifluoressigsäure der	

	Konzentrationsklasse 2 ($> 0,5 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).128
Abbildung 68:	Desphenyl-Chloridazon: Die Verteilung der Anbaufläche der Hauptkultur (Zuckerrübe) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (mitte, grün) und Überschätzung (rechts, gelb) als Violinenplot dargestellt.129
Abbildung 69:	Metazachlor ESA: Die Verteilung der Anbaufläche der Hauptkultur (Winterraps) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Überschätzung (rechts, gelb) als Violinenplot dargestellt.129
Abbildung 70:	Metolachlor OXA: Die Verteilung der Anbaufläche der Hauptkultur (Mais) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Überschätzung (rechts, gelb) als Violinenplot dargestellt.130
Abbildung 71:	Metolachlor NOA 413173: Die Verteilung der Tiefe der Filteroberkante wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Überschätzung (rechts, gelb) als Violinenplot dargestellt.130
Abbildung 72:	Metazachlorsäure: Die Verteilung der Tiefe der Filteroberkante wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Überschätzung (rechts, gelb) als Violinenplot dargestellt.....131
Abbildung 73:	Stärke der Feature-Interaktionen für vier ausgewählte Metabolite.....132
Abbildung 74:	Stärke der Feature-Interaktionen für vier ausgewählte Metabolite.....132
Abbildung 75:	Vergleich der Performance der zwei Algorithmen (Neuronale Netze und Random-Forest) anhand der Abweichung der tatsächlichen und vorhergesagten Konzentrationsklassen für die binären Modelle der reduzierten Feature-Auswahl.133
Abbildung 76:	Modellgüte der Random-Forest Modelle für die erweiterte Feature-Auswahl für zwei Chloridazon- und zwei Metazachlor-Metaboliten.....135
Abbildung 77:	Modellgüte der Random-Forest Modelle für die erweiterte Feature-Auswahl für drei Metolachlor-Metaboliten und die Trifluoressigsäure.135
Abbildung 78:	Die jeweils 15 wichtigsten Features nach der Permutation Feature Importance für das binäre Random-Forest Modell der

	erweiterten Feature-Auswahl für Desphenyl-Chloridazon, Methyl-desphenyl-Chloridazon, Metazachlor ESA und Metazachlorsäure.	137
Abbildung 79:	Fortführung von Abbildung 78 für Metolachlor ESA, Metolachlor OXA, Metolachlor NOA 413173 und die Trifluoressigsäure.	138
Abbildung 80:	Feature-Effekte von Desphenyl-Chloridazon im binären Random-Forest Modell der erweiterten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.	139
Abbildung 81:	Feature-Effekte von Methyl-desphenyl-Chloridazon, Fortführung vorheriger Abbildungen.	139
Abbildung 82:	Feature-Effekte von Metazachlor ESA, Fortführung vorheriger Abbildungen.	140
Abbildung 83:	Feature-Effekte der Metazachlorsäure, Fortführung vorheriger Abbildungen.	140
Abbildung 84:	Feature-Effekte von Metolachlor ESA, Fortführung vorheriger Abbildungen.	140
Abbildung 85:	Feature-Effekte von Metolachlor OXA, Fortführung vorheriger Abbildungen.	141
Abbildung 86:	Feature-Effekte von NOA 413173, Fortführung vorheriger Abbildungen.	141
Abbildung 87:	Feature-Effekte von TFA, Fortführung vorheriger Abbildungen.	141
Abbildung 88:	Die Modellgüte der binären Random-Forest-Modelle der reduzierten Feature-Auswahl für drei anhand der Mobilität eingeteilte Wirkstoffgruppen.	143
Abbildung 89:	Permutation-Importance für die drei Wirkstoffgruppenmodelle nach Mobilität (reduzierte Feature-Auswahl).	145
Abbildung 90:	Permutation, Gini impurity und korrigierte Gini impurity exklusiv für die moderat mobilen Wirkstoffgruppenmodelle.	145
Abbildung 91:	Feature-Effekte der moderat mobilen Wirkstoffgruppe im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.	146
Abbildung 92:	Tatsächliche und vorhergesagte Konzentrationsklasse exklusiv für moderat mobile Wirkstoffgruppen aller Grundwassermessstellen.	147
Abbildung 93:	Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen exklusiv für moderat mobile Wirkstoffe für die Konzentrationsklasse 2 ($> 0,01 \mu\text{g L}^{-1}$; mitte),	

	wo das Ergebniss unterschätzt wird (rechts) und wo es überschätzt wird (links).....	148
Abbildung 94:	Feature-Importance gemäß Permutation für die Wirkstoffgruppen nach Mobilität (erweitere Feature-Auswahl).	149
Abbildung 95:	Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der mobilen und sehr mobilen Wirkstoffe mit erweiterter Feature-Auswahl.	150
Abbildung 96:	Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der moderat mobilen Wirkstoffe mit erweiterter Feature-Auswahl.....	150
Abbildung 97:	Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der wenig mobilen Wirkstoffe mit erweiterter Feature-Auswahl.....	151
Abbildung 98:	Anzahl der Messstellen (links) und prozentuale Anteil an Messstellen (rechts) mit und ohne Überschreitungen des 0,05 $\mu\text{g L}^{-1}$ Konzentrationswertes durch die Wirkstoffe Desphenyl-Chloridazon, Metazachlor ESA und Metolachlor ESA in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.	164
Abbildung 99:	Der prozentuale Anteil an Messstellen mit und ohne Überschreitungen des 0,1 $\mu\text{g L}^{-1}$ Grenzwertes durch Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung wird dargestellt.	165
Abbildung 100:	Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des 0,01 $\mu\text{g L}^{-1}$ Konzentrationswertes durch sehr mobile und mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.	165
Abbildung 101:	Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des 0,01 $\mu\text{g L}^{-1}$ Konzentrationswertes durch wenig mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.	166
Abbildung 102:	Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des 0,01 $\mu\text{g L}^{-1}$ Konzentrationswertes durch moderat mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.	166
Abbildung 103:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von 0,1 $\mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO	

	DE L-Modell entnommen wurden, für den Wirkstoff Bentazon.	169
Abbildung 104:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,1 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Bentazon.	170
Abbildung 105:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Bentazon.	170
Abbildung 106:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Bentazon.	171
Abbildung 107:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Metazachlor.	171
Abbildung 108:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Metazachlor.	172
Abbildung 109:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Metolachlor.....	172
Abbildung 110:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Metolachlor.	173
Abbildung 111:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Terbuthylazin.....	173

Abbildung 112:	Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Terbutylazin.....	174
Abbildung 113:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metazachlor ESA.	174
Abbildung 114:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metazachlor ESA.	175
Abbildung 115:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metazachlorsäure.	175
Abbildung 116:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metazachlorsäure.	176
Abbildung 117:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metolachlor ESA.	176
Abbildung 118:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metolachlor ESA.	177
Abbildung 119:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metolachlor OXA.....	177
Abbildung 120:	Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metolachlor OXA.	178

Abbildung 121:	Kategorien von Messstellen aus dem FARM-Projekt die für das Nachzulassungsmonitoring (NZM) relevant oder nicht relevant sind.	189
Abbildung 122:	Lage der NZM-kompatiblen Messstellen des FARM-Projektes differenziert nach strikten (A) und toleranten (B) Kriterien...	190
Abbildung 123:	Statistische Kennzahlen zum mittleren (A) und maximalen Grundwasserflurabstand (B) je Bundesland.....	190
Abbildung 124:	Statistische Kennzahlen der Metriken zur Filterunterkante (A), Filteroberkante (B) und Filterlänge (C) je Bundesland.	191
Abbildung 125:	Informationen zur Messhäufigkeit des Grundwasserflurabstandes je Messstelle und Bundesland (A) und Jahr der letzten Messung, in der die Messung durchgeführt wurde (B).....	192
Abbildung 126:	Häufigkeit der dominierenden Landnutzung innerhalb des Puffers einer NZM-kompatiblen Messstelle abgeleitet aus Sentinel-Daten differenziert für die strikte und tolerante Kategorie.	192
Abbildung 127:	Übersicht der im Projekt verfügbaren InVeKoS-Daten pro Jahr und Bundesland (grau = Daten verfügbar).	214
Abbildung 128:	Beispiel von verarbeiteten InVeKoS-Daten, die den im Projekt verwendeten Kulturklassen zugeordnet und in einem 1-km-Radius um die Grundwassermessstelle ausgeschnitten wurden.	215
Abbildung 129:	Rechenbeispiel für die klassenspezifische Ermittlung von Precision, Recall und F1-Score sowie die Berechnung des Macro-F1-Scores für eine binäre Klassifikation (2 Klassen). Für die Berechnungen werden vorhergesagte und tatsächliche Klasse zunächst in <i>One-vs-rest</i> -Konfusionsmatrizen umgewandelt. Erklärungen und Formeln zu den Berechnungen finden sich in 2.2.3.3.	216
Abbildung 130:	Rechenbeispiel für die klassenspezifische Ermittlung von Precision, Recall und F1-Score sowie die Berechnung des Macro-Scores für eine Mehrklassen-Klassifikation (3 Klassen). Für die Berechnungen werden vorhergesagte und tatsächliche Klasse zunächst in <i>One-vs-rest</i> -Konfusionsmatrizen umgewandelt. Erklärungen und Formeln zu den Berechnungen finden sich in 2.2.3.3.	217
Abbildung 131:	Partial-Dependence-Plots für Desphenyl-Chloridazon.....	219
Abbildung 132:	Partial-Dependence-Plots für Methyldesphenyl-Chloridazon.	220
Abbildung 133:	Partial-Dependence-Plots für die Metazachlorsäure.	220
Abbildung 134:	Partial-Dependence-Plots für Metazachlor ESA.	221
Abbildung 135:	Partial-Dependence-Plots für Metolachlor ESA.	221

Abbildung 136:	Partial-Dependence-Plots für Metolachlor OXA.	222
Abbildung 137:	Partial-Dependence-Plots für Metolachlor Metabolit NOA 413173.	222
Abbildung 138:	Partial-Dependence-Plots für Trifluoressigsäure.	223

Tabellenverzeichnis

Tabelle 1:	Übersicht der Persistenz- und Mobilitätsklassen anhand des Adsorptionskoeffizienten (K_{Foc}) und der Abbauzeit im Boden (DT_{50}).	42
Tabelle 2:	Bei Klassifikationsproblemen können bei Abgleich der vorhergesagten mit der tatsächlichen Klasse die vier möglichen Fälle TP, FN, FP und TN dargestellt werden.	52
Tabelle 3:	Wirkstoffe und Metaboliten, die in allen dreizehn Flächenbundesländern gesucht werden, sind mit ihrer Pestizidklasse und ihrer Zulassungssituation gelistet.....	76
Tabelle 4:	Ranking der Wirkstoffe und Metaboliten nach Anzahl der Messstellen, an denen $0,1 \mu\text{g L}^{-1}$ überschritten wurden. Stand der Relevanz der Metaboliten ist März 2023, seither wurden zahlreiche xM als rM eingestuft.	86
Tabelle 5:	Modellgüte (Macro-F1-Score) der Logistischen Regression für die binären Metaboliten-Modelle der reduzierten Feature- Auswahl sowie der Neuronalen Netze für die reduzierte und erweiterte Feature-Auswahl.	97
Tabelle 6:	Kriterien für die Auswahl von potentiellen Messstellen für das Nachzulassungsmonitoring (NZM).	188
Tabelle 7:	Zusammenfassung der Messstellen die für die Vulnerabilitätsbewertung des Grundwassers durch Pflanzenschutzmittel im FARM-Projekt durch die Länder bereitgestellt und ausgewertet wurden und Anzahl der Messstellen die für die Datennachnutzung veröffentlicht werden dürfen.....	209
Tabelle 8:	Übersicht zu der reduzierten und erweiterten Auswahl der Eingangsdaten (features) die für das maschinelle Lernen verwendet wurden.....	210
Tabelle 9:	Methodenübersicht zu Verfahren der Vulnerabilitätsbewertung die in der Grundwasserbewertung angewendet wurden.	224

Abkürzungsverzeichnis

Abkürzung	Erläuterung
AF	Attenuation factor
AVI	Aquifer vulnerability index
BfR	Bundesinstitut für Risikobewertung
BGR	Bundesanstalt für Geowissenschaften und Rohstoffe
BÜK	Bodenübersichtskarte
BVL	Bundesamt für Verbraucherschutz und Lebensmittelsicherheit
DB	Datenbank
DLM	Digitales Landschaftsmodell
DOC	Gelöster organischer Kohlenstoff
DT₅₀	Abbauzeit im Boden wo 50% eines Wirkstoffes abgebaut sind
ECHA	Europäische Chemikalienagentur
EFSA	Europäischen Lebensmittelbehörde
EU	Europäische Union
FOK	Filteroberkante
FUK	Filterunterkante
GOK	Geländeoberkante
GOW	Gesundheitlicher Orientierungswert
GW	Grundwasser
GWL	Grundwasserleiter
GWN	Grundwasserneubildung
HÜK	Hydrologische Übersichtskarte
HYRAUM	Hydrogeologischer Großraum
InVeKoS	Integriertes Verwaltungs- und Kontrollsystem
JKI	Julius-Kühn-Institut
KA5	Bodenkundliche Kartieranleitung
K_{Foc}	Adsorptionskoeffizienten
KM	Kaplan-Maier-Schätzer
LAWA	Bund/Länder-Arbeitsgemeinschaft Wasser
LI	Leaching Index
LOD	Limit of detection
LOQ	Limit of quantification

Abkürzung	Erläuterung
MLE	Maximum-Likelihood-Estimation
MST	Messstelle
NN	Neural network
nrM	Nicht-relevante Metaboliten
NZM	Nachzulassungsmonitoring
PAPA	Panel Pflanzenschutzmittel-Anwendungen
PDP	Partial dependence plots
PPDB	Pesticide Property Data Base
PSM	Pflanzenschutzmittel
RF	Random-Forest
rM	Relevante Metaboliten
RNN	Recurrent neural network
ROS	Regression-on-Order-Statistics
SGD	Staatliche Geologische Dienste
SWR	Sickerwasserrate
TrinkwEGV	Trinkwassereinzugsgebietsverordnung
UBA	Umweltbundesamt
WRRL	Wasserrahmenrichtlinie
WVU	Wasserversorgungsunternehmen

Zusammenfassung

Hintergrund und Zielsetzung

Im Zulassungsverfahren für Pflanzenschutzmittel (PSM) gewinnen zielgerichtete Monitoringstudien als höherwertige Studien in der gestuften Risikobewertung Grundwasser zunehmend an Bedeutung. Einschätzungen zur Vulnerabilität von Standorten und Messstellen gegenüber Einträgen von PSM-Wirkstoffen und ihren Abbauprodukten ins Grundwasser sind dafür im Rahmen der harmonisierten EU-weiten Bewertung essentiell. Eines der Ziele des Projektes war es zu evaluieren, welche Faktoren die Vulnerabilität des Grundwassers gegenüber dem Eintrag von Wirkstoffen und Metaboliten von PSM in Deutschland maßgeblich bestimmen. Darauf aufbauend sollte die Frage beantwortet werden, wie sich die Vulnerabilität von Grundwasser gegenüber PSM-Einträgen quantifizieren, klassifizieren und räumlich vorhersagen lässt. In einem weiteren Schritt sollte geprüft werden, ob die Modellergebnisse in die Entwicklung eines bundesweiten Indexverfahrens zur Vulnerabilität des Grundwassers gegenüber PSM überführt werden können. Damit verbunden ist die Frage nach neuen Erkenntnissen zu relevanten Standorteigenschaften, welche die Auswahl von Messstellen für zielgerichtete Nachzulassungsmonitoring (NZM) -studien beeinflussen könnten und welche Empfehlungen sich daraus ableiten lassen.

Datensatz und Fundsituation

Um die Fragen zu beantworten wurden in diesem Projekt Grundwasserqualitätsdaten der dreizehn Flächenbundesländer für eine Vielzahl an Wirkstoffen und Metaboliten von PSM in einer umfangreichen Datenbank zusammengeführt, harmonisiert und ausgewertet. Die Datenbank wurde mit Informationen zu den Stoff- und Standorteigenschaften aus anderen Datenquellen verknüpft. Insgesamt sind über 26.000 Messstellen und über 500 Stoffe in der Datenbank enthalten. Die älteste Messung ist aus dem Jahr 1973, die größte Messstellendichte und das breiteste Stoffspektrum stellt sich jedoch erst ab etwa 2010 ein. Nachdem vorwiegend nach Wirkstoffen gesucht wurde, werden ab 2007 vermehrt Metaboliten in das Untersuchungsprogramm aufgenommen. Die Monitoringstrategien, d.h. Anzahl und Art der Stoffe, nach denen gesucht wird, sowie Messintervalle und räumliche Abdeckung unterscheiden sich deutlich je nach Bundesland und über die Zeit. Die meisten untersuchten Stoffe sind Herbizide, gefolgt von Insektiziden und Fungiziden sowie weiteren Stoffklassen. Die untersuchten Metaboliten sind größtenteils Metaboliten von Herbiziden.

Betrachten wir die gesamte Datenbank, dann wurden im Gesamtzeitraum Konzentrationen oberhalb der jeweiligen Bestimmungsgrenze (LOQ) mindestens einmal für 315 verschiedene PSM-Wirkstoffe und Metaboliten an insgesamt 15.301 Messstellen in allen dreizehn Flächenbundesländern gemessen. Das heißt, dass an 58,4% der Messstellen im Datensatz Wirkstoffe und Metaboliten nachgewiesen werden konnten. Gleichzeitig liegen 96,6 % aller im Datensatz enthaltenen Messungen (unabhängig vom Stoff oder der Messstelle) unterhalb der LOQ. Überschreitungen von $0,1 \mu\text{g L}^{-1}$ von insgesamt 156 Wirkstoffen und relevanten Metaboliten finden sich im Gesamtdatensatz an 6.256 Messstellen (19,9%) in den Bundesländern. Im Zeitraum ab 2010 sind es 129 Wirkstoffe und relevante Metaboliten, die an insgesamt 4.377 Messstellen (23,9% der ab 2010 beprobten Messstellen) den Grenzwert überschreiten.

In Cooke et al. (2024) wurde der Anteil der Messungen in den fünf Konzentrationsklassen ($< \text{LOD}$, $< \text{LOQ}$, $< 0,1$, $> 0,1$ und < 1 , sowie $> 1 \mu\text{g L}^{-1}$) jeweils für Wirkstoffe und relevante Metaboliten bzw. für nicht-relevante Metaboliten anhand ausgewählter Standortfaktoren und in zwei Zeiträumen untersucht. Hierbei wurde die Verteilung der Konzentrationsklassen hinsichtlich vier Aquifer-Hohlraumarten (Poren-, Kluft-, Kluft-/Poren- und Kluft-/Karst-

Grundwasserleiter), drei Landnutzungsklassen (Ackerland/Grünland, Forst und Bebauung) sowie der Tiefe der Filteroberkante unterhalb der Geländeoberkante in fünf Tiefenbereichen (0-2, 2-10, 10-20, 20-50, 50-100 m) dargestellt. Für den Zeitraum 2010-2019 lässt sich zusammenfassend beobachten, dass höhere Anteile von höheren Konzentrationsbereichen in allen Grundwasserleitern vorkommen, allerdings gehäuft in Porengrundwasserleitern. Weiterhin verweisen die Ergebnisse auf den Einfluss der Landnutzung in der Umgebung der Messstelle (anteilig mehr höhere Konzentrationsklassen bei Ackerland) und der Filtertiefe (anteilig mehr höhere Konzentrationsklassen bei eher oberflächennahen Filtern). Überschreitungen von $0,1 \mu\text{g L}^{-1}$ zeigten nur für wenige Stoffe (vorwiegend Metaboliten) bivariate Korrelationen mit Standortfaktoren mit einem Koeffizienten $> 0,3$. Auch die Anzahl an Standortfaktoren, die überhaupt Korrelationen $> 0,3$ aufweisen, ist reduziert. Hierbei fallen wie zu erwarten positive Korrelationen mit den zugehörigen und weiteren, ggf. der Fruchtfolge geschuldeten, Kulturen sowie der mittleren Nitrat- und Kaliumkonzentration im Grundwasser auf.

Methoden des Maschinellen Lernens

Das maschinelle Lernen umfasst statistische Methoden, bei denen Computeralgorithmen aus Daten Modelle entwickeln, die u.a. zur Vorhersage genutzt werden können. Die relevanten Regeln und Zusammenhänge werden dabei nicht vorgegeben, sondern explizit aus den Daten entnommen. Der Lernprozess wird als Training bezeichnet und beinhaltet ein iteratives Anpassen von Modellparametern sowie den Abgleich von Modellvorhersagen anhand der Ausgangsdaten, bei der die Abweichung zwischen Daten und Vorhersage sukzessiv verringert wird. In diesem Projekt wurden zwei verschiedene Algorithmen – Random Forest und Neuronale Netze – in einer Klassifikation angewandt. Zuvor wurde die Zielvariable und die erklärenden Variablen (sog. Features) definiert bzw. zusammengestellt. Die Gewichtung und Analyse des Einflusses der Features auf die Vorhersage der Zielvariable waren Kernanliegen und wurde mittels entsprechender Methoden berechnet.

Die unterschiedliche und z.T. unzureichende zeitliche Auflösung der Messdaten und erklärenden Variablen führt dazu, dass ein räumlicher, zeitlich aggregierter Modellansatz gewählt wurde. Die stark links-zensierten Messdaten der Wirkstoffe und Metaboliten erschwerten die direkte Verwendung der Stoffkonzentrationen als Zielvariable und hatten einen entscheidenden Einfluss auf die Auswahl an Stoffen, für die Modelle gerechnet werden konnten. Daher wurden die Stoffmessungen in Konzentrationsklassen eingeteilt und diese an den Messstellen pro Stoff vorhergesagt (Klassifikation). Bei den binären Konzentrationsklassen handelt es sich um Konzentrationsklasse 1) $\leq 0,05 \mu\text{g L}^{-1}$ sowie eine Konzentrationsklasse 2) $> 0,05 \mu\text{g L}^{-1}$. Eine für robuste Modelle ausreichende Datenlage (ausreichend Werte oberhalb der Bestimmungsgrenze) lag für acht Metaboliten vor. Es wurden folglich binäre Klassifikationsmodelle mittels zweier Architekturen für die Metaboliten Desphenyl-Chloridazon, Methylphenyl-Chloridazon, Metolachlor ESA, Metolachlor OXA und NOA 413173 sowie Metazachlor ESA, Metazachlorsäure und Trifluoressigsäure (TFA) trainiert. Bei diesen Metaboliten handelt es sich um vorwiegend relativ mobile und persistente Abbauprodukte von Wirkstoffen, bei denen eine lange Anwendungshistorie bekannt ist. Als Features (reduzierte Feature-Auswahl) wurden die Fläche der jeweiligen Haupt- bzw. Nebenkulturen im 1 km-Radius um die Messstelle, die Tiefe der Filter-Oberkante, die mittlere Grundwasserneubildung, der Sandgehalt bzw. der Gehalt an organischem Kohlenstoff im Boden, die Makroporenklasse sowie die hydraulische Leitfähigkeit bzw. die Hohlraumart des Aquifers verwendet.

In einer sog. erweiterten Feature-Auswahl wurden zusätzlich zu den genannten Features auch der hydrogeologische Großraum, langjährige Features zur Lufttemperatur und zum Niederschlag, die Geländehöhe, Hangneigung, die Gesteinsart und der geochemische Typ des

Aquifers, mehrere Features zur Verortung der Messstelle im oberirdischen Einzugsgebiet, Schluff- und Tongehalte, die Größe der Ackerfläche, sowie die langjährige mittlere Sickerwasserrate verwendet. Weiterhin wurden Features, die aus den grundwasserchemischen Messwerten an den Messstellen abgeleitet wurden, und als zusätzliche Zeiger der Grundwasserqualität gesehen werden können, ebenfalls verwendet (Nitrat, Ammonium, Magnesium, Sulfat, Chlorid, Kalium und Natrium-Konzentrationen). Die Berücksichtigung von Wirkstoffen gelang aufgrund des hohen Anteils links-zensierter Messungen nur begrenzt. Die reduzierte und die erweiterte Feature-Auswahl wurde für einzelne Wirkstoffe (Bentazon, Atrazin und Simazin) sowie für drei, nach ihrer Mobilität (K_{Foc} -Werte) zusammengefasste Wirkstoffgruppen (sehr mobil und mobil; moderat mobil, wenig mobil) gerechnet.

Ergebnisse des Maschinellen Lernens

Metaboliten

Die binären Metaboliten-Modelle der reduzierten Feature-Auswahl zeigen größtenteils Modellgüten, die signifikant höher sind als Zufallszuordnungen (Macro-F1-Score > 0,7) und die Mehrheit der Messstellen (~80%) konnten korrekt vorhergesagt werden. Somit konnten mit einer überschaubaren Anzahl von räumlichen und zeitlich-aggregierten Features räumliche Modelle für acht Metaboliten mit guter Vorhersagegüte trainiert werden.

Unabhängig vom Algorithmus wurde bei fast allen Metaboliten die jeweilige Hauptkultur als das mit Abstand wichtigste Feature identifiziert. Als weitere, nachfolgende Features sind die Filter-Oberkante, der Sandgehalt, die Grundwasserneubildung und z.T. auch Nebenkulturen zu nennen. Die Reihenfolge dieser Features kann sich je nach Methode etwas unterscheiden. Bei TFA wird die Filter-Oberkante und der Sandgehalt als am wichtigsten eingestuft, es folgen zahlreiche Kulturen. Den hydrogeologischen Features und der Makroporenklasse wurde für alle Stoffe eine geringe Feature-Gewichtung zugeordnet. Bei allen Stoffen findet sich der Zusammenhang, dass eine größere Fläche der Hauptkultur mit der höheren Konzentrationsklasse und eine tiefere Filter-Oberkante mit der niedrigeren Konzentrationsklasse assoziiert ist. Bei den Features Sandgehalt und Grundwasserneubildung zeigen Unterschiede zwischen den Stoffen, was zum Teil mit der Verteilung dieser Features in den jeweiligen Anbaugebieten in Zusammenhang gebracht werden kann. Wir gehen davon aus, dass die Makroporenklasse - ähnlich dem Sandgehalt, nur umgekehrt - auf die Verteilung der Bodeneigenschaften in Anbaugebieten hindeutet bzw. komplementär zum Feature Anbaufläche auf Gebiete mit PSM-Einsatz hindeutet.

Die Hinzunahme weiterer Features führte bei allen Stoffen zu einer Reduzierung der verfügbaren Messstellenanzahl. Die Modellgüten waren dennoch höher als bei der reduzierten Feature-Auswahl. Die jeweilige Hauptkultur ist bei fast allen Stoffen nach wie vor das wichtigste Feature. Ebenfalls sehr relevant sind die mittleren Gehalte an Nitrat, Sulfat und Kalium im Grundwasser. Diese Features weisen einen mit der Anbaukultur vergleichbaren Einfluss auf die Vorhersage der Konzentrationsklassen auf und sind wahrscheinlich ebenfalls als Zeiger für landwirtschaftlich genutzte Gebiete zu interpretieren. Die grundwasserchemischen Features weisen eine mit der Landnutzung vergleichbaren Einfluss auf die Stoffkonzentrationen auf. Bei den meisten Metaboliten ist der Nitratgehalt im Grundwasser in Anbaugebieten höher, daher kann vermutet werden, dass die grundwasserchemischen Features mit dem Anbau der Hauptkultur in einem räumlichen Zusammenhang stehen. In Kombination mit der Analyse der Feature-Effekte lässt sich ableiten, dass diese grundwasserchemischen Parameter als Zeiger für landwirtschaftliche Aktivität und ggf. als ein Hinweis auf die Intensität dieser gesehen werden können. Zwar kann das Vorkommen von Nitrat auch ein Hinweis auf ein erhöhtes Versickerungspotenzial sein, in diesem Kontext ist jedoch davon auszugehen, dass die Modelle Zusammenhänge vorwiegend aufgrund der räumlichen Assoziation lernen. Weitere Features

(langfristige meteorologische Bedingungen, Höhenlage etc.) können aufgrund ihrer Wechselbeziehung (Multi-Kollinearität) ebenfalls auf landwirtschaftliche Gebiete verweisen.

Wirkstoffe

Die Vorhersagegüte der mobilen und sehr mobilen bzw. wenig mobilen Wirkstoffgruppenmodelle (reduzierte Feature-Auswahl) sind hinsichtlich der Vorhersage der höheren Konzentrationsklasse nicht robust genug, daher ist die Feature-Importance dieser Modelle mit Unsicherheiten behaftet. Die Modelle der mobilen und sehr mobilen Wirkstoffe zeigen jedoch eine Feature-Importance auf, die ähnlich zu den acht Metaboliten ist. Die moderat mobilen Wirkstoffe weisen den hydrogeologischen und Boden-Features eine höhere Gewichtung zu als dem Feature Ackerland. Bei den Modellen der erweiterten Feature-Auswahl wurde bei allen Stoffgruppen und den Einzelstoffen eine leicht bessere Modellgüte erreicht. Der Nitratgehalt ist wichtigstes Feature bei den moderat mobilen und wenig mobilen Wirkstoffen.

Einordnung der Ergebnisse im Kontext der Vulnerabilität und Abgleich mit weiteren Vulnerabilitätsverfahren

Die in den Grundwasserproben gefundenen Konzentrationen können das Ergebnis mehrerer sich überschneidender Prozesse sein. Anwendungsdaten zur tatsächlichen Applikation von PSM-Wirkstoffen sind das wichtigste Bindeglied zur Plausibilisierung von Landnutzung und Stoffkonzentrationen im Grundwasser, standen jedoch nicht zur Verfügung. Die Quantifizierung von Vulnerabilität (stark, mäßig, schwach) ist vom Einfluss der tatsächlichen Anwendung schwer zu entkoppeln, wenn zu letzterer keine langjährigen Daten vorliegen. Auch keine nachweisbaren oder sehr geringen Konzentrationen sind folglich nicht zwangsläufig gleichbedeutend mit geringer intrinsischer Vulnerabilität, da der Stoff ggf. nicht ausgebracht wurde. Aufgrund dieser Limitation sind der Bewertung der Grundwasservulnerabilität aus den vorliegenden Messdaten Grenzen gesetzt, insbesondere im Hinblick auf die Entwicklung eines Vulnerabilitätsindizes aus diesen Beobachtungsdaten.

Letztendlich wird geschlussfolgert, dass sich anhand der vorliegenden Datengrundlage kein robuster Vulnerabilitätsindex für PSM entwickeln lässt. Daher sollte im nächsten Schritt geprüft werden, ob existierende Vulnerabilitätskonzepte anhand von Monitoringdaten evaluiert werden könnten. Zum einen wurde der Abgleich zwischen Schutzfunktion der Grundwasserüberdeckung (nach Hölting et al., 1995 und LAWA) und den Monitoringdaten und zum anderen zwischen GeoPELMO DE-Simulationsergebnissen und den Monitoringdaten durchgeführt. Die simulierten Sickerwasserkonzentrationen wurden in vier räumliche Perzentilklassen unterteilt.

Der Abgleich zwischen der Schutzfunktion der Grundwasserüberdeckung und den Konzentrationsklassen von Metaboliten an den Messstellen im FARM-Datensatz zeigte keinen Zusammenhang. Allein bei den Wirkstoffen zeigte sich eine Tendenz, dass Standorte mit „ungünstiger“ Schutzfunktion anteilig mehr Messstellen mit Funden aufweisen. Gleichzeitig kommen an Standorten aller drei Kategorien der Schutzfunktion ein gewisser Prozentsatz an Messstellen mit Funden vor. In der Kategorie „günstig“ sind Standorte vertreten, bei denen geschätzt wird, dass es mehr als 10 und z.T. mehr als 25 Jahre dauern kann, bis ein Stoff das Grundwasser erreicht. Bei den betrachteten Metaboliten handelt es sich ausschließlich um mobile und sehr mobile Metaboliten, die zum Teil von Wirkstoffen mit jahrzehntelanger Anwendungshistorie stammen. Es wird vermutet, dass die Beschreibung der intrinsischen Vulnerabilität (nach Hölting et al. 1995 und der LAWA) für diese Metaboliten mit hoher Mobilität, die von Altstoffen stammen, nicht hinreichend geeignet ist. Die Abbauprodukte haben mittlerweile zahlreiche Grundwasserleiter unterschiedlicher Kategorien erreicht und kommen ubiquitär vor.

Die Abgleiche zwischen Monitoringdaten und den prozessorientierten Modellergebnissen mit GeoPELMO DE zeigten größtenteils keine Übereinstimmungen. Grundsätzlich lässt sich beobachten, dass sich bei allen betrachteten Wirkstoffen und Metaboliten Messstellen mit entsprechenden Funden in allen räumlichen Perzentilklassen der Modellvorhersagen finden. Für Metazachlor ESA findet sich bei GeoPELMO DE (Version FOCUS L; Leaching, d.h. Versickerung durch Matrixfluss) der erwartete Zusammenhang (anteilig mehr Überschätzungen in der Perzentilkategorie 75-100), während bei GeoPELMO DE (Version FOCUS L+R+M+D; neben Leaching werden die Prozesse Oberflächenabfluss, Makroporenfluss und Drainage berücksichtigt) dies nicht klar ersichtlich ist. Für Metazachlorsäure zeigt sich bei beiden Modellkonfigurationen eine Tendenz zu anteilig mehr Überschätzungen in der Perzentilkategorie 75-100. Für Metolachlor ESA und Metolachlor OXA findet sich jedoch kein sichtbarer Zusammenhang. Unter den Wirkstoffen findet sich bei Bentazon ein gegenläufiger Zusammenhang mit einem höheren Anteil der Messstellen mit Funden in der Perzentilkategorie 0-25 in beiden Modellkonfigurationen. Im Falle von Metolachlor, Metazachlor und Terbuthylazin findet sich kein klarer Zusammenhang. Eine leichte Tendenz zu einem erwartbaren Zusammenhang kann höchstens bei Terbuthylazin für die Modellkonfiguration GeoPELMO DE (FOCUS L+R+M+D) diskutiert werden.

Insgesamt lässt sich ableiten, dass dieser Abgleich für die meisten Stoffe (sieben Metaboliten und vier Wirkstoffe) keine Übereinstimmung zwischen dem Anteil der Messstellen mit Funden und der höheren räumlichen Perzentilkategorie gezeigt hat. Es werden Messstellen mit den jeweiligen Konzentrationsüberschreitungen in allen räumlichen Perzentilklassen gefunden. Einen Vorteil von GeoPELMO DE (FOCUS L+R+M+D) mit zusätzlichen Prozessen wie Oberflächenabfluss, Makroporenfluss und Drainage gegenüber der FOCUS-basierten Version GeoPELMO DE (FOCUS L) lässt sich hieraus nicht ableiten. Bei diesem Ergebnis sollte berücksichtigt werden, dass die Entwickler von GeoPELMO DE einen begrenzten Einfluss der zusätzlichen Prozesse Oberflächenabfluss, Makroporenfluss und Drainage auf die modellgestützte Vorhersage von Sickerwasserkonzentrationen für Metaboliten postulieren. Das könnte der Grund für die vergleichsweise geringen Unterschiede in unseren Vergleichen der Monitoringdaten mit beiden Modellversionen für die untersuchten Metaboliten sein.

Zusammenfassend lässt sich sagen, dass die Überlegungen zur Einordnung der ML-Modellergebnisse hinsichtlich ihrer Aussagekraft in der Vulnerabilitätsbewertung sowie der Abgleich der index- und der prozessbasierten Verfahren mit den Monitoringdaten zahlreiche Fragen aufwerfen. Beim betrachteten Index-Verfahren steht die Vermutung im Raum, dass diese bei Stoffen mit langer Anwendungshistorie und hoher Mobilität unzureichend sind. Die Abweichung zwischen den Simulationen aus Versickerungsmodellen und den Monitoringdaten zeigt in erster Linie die Skalenfrage auf: Sind die simulierten Sickerwasserkonzentrationen im Boden und die Konzentrationen im Grundwasser zeitlich und räumlich vergleichbar? Aus diesen Gründen lassen sich aus diesem Projekt keine klaren Empfehlungen ableiten, welches der beiden zur Verfügung stehenden Verfahren besser geeignet ist, bspw. um vulnerable Standorte für NZM-Studien zu identifizieren.

Die ML-Modelle für mobile und sehr mobile Metaboliten bzw. Wirkstoffe neigen dazu, die landwirtschaftliche Nutzfläche (konkrete Kultur oder Ackerland allgemein) als eines der wichtigsten, wenn nicht als das wichtigste Feature zu identifizieren. Hydrogeologische Eigenschaften spielen eine deutlich untergeordnete Rolle. Diese Beobachtung deckt sich mit der Abweichung zu den Schutzfunktionskategorien.

Bei den Wirkstoff- bzw. Wirkstoffgruppenmodellen war dies nicht immer der Fall. Gerade bei den moderat mobilen Wirkstoffen zeigten sich andere Features (Boden, Hydrogeologie) auf den oberen Rangplätzen. Der Abgleich zwischen Hölting und den Funden der Wirkstoffe zeigte in der Tendenz, dass die Schutzfunktion der Grundwasserüberdeckung Wirkstofffunde besser

beschreiben kann, bzw. dass die Wirkstoffe sich zur räumlichen Differenzierung von intrinsischer Vulnerabilität ggf. besser eignen könnten als die untersuchten Metaboliten mit ihren Eigenschaften. Welche konkrete Rolle die Mobilität der Stoffe spielt, konnte nicht abschließend geklärt werden. Dies wird durch die ähnliche Feature-Gewichtung zwischen Metaboliten und mobilen Wirkstoffen unterstrichen. Auch die mobilen und sehr mobilen Wirkstoffmodelle zeigten eine geringe Feature-Gewichtung der hydrogeologischen Features während das Feature Ackerland als relativ wichtig eingestuft wurde. Selbstverständlich ist beim Vergleich zwischen Wirkstoffen und Metaboliten Vorsicht geboten, da letztere erst im Boden entstehen und dort ggf. länger gespeichert werden können. Wir vermuten, dass ggf. aus weniger mobilen Stoffen mehr über die intrinsische Vulnerabilität gelernt werden könnte. Gleichzeitig liegt es in der Natur der Sache, dass solche Stoffe und gerade Wirkstoffe in geringeren Konzentrationen auftauchen bzw. einen derart hohen Anteil zensierter Messungen aufweisen, dass daraus keine aussagekräftigen Abgleiche, geschweige denn robuste Modelle, trainiert werden können. Eine Verringerung und Harmonisierung der Bestimmungsgrenzen könnte dazu beitragen, bundesweite Wirkstofffunde zukünftig besser auswerten zu können.

Empfehlungen für das Nachzulassungsmonitoring

Aus den Ergebnissen der harmonisierten Datenbank lässt sich eine Vorauswahl an Messstellen treffen, die für Nachzulassungsmonitoring- (NZM) Studien in Deutschland als potenziell geeignet betrachtet werden. Von den ca. 26.000 Messstellen, die im FARM Projekt untersucht wurden, sind 7.675 von den definierten Kriterien nicht NZM-kompatibel. Gründe dafür sind grundwasserferne Standorte oder Filteroberkanten, die nicht in der Nähe der Grundwasseroberfläche zu verorten sind. Grundwasserstandsdaten (n = 6.377) fehlten gegenüber Stammdaten der Messstelle (n = 1.407) besonders häufig und ermöglichten keine genaue Zuweisung dieser Standorte. Beide Informationen zum Grundwasserstand und Stammdaten der Messstelle fehlten ebenfalls besonders häufig (n = 6.686). Bei 3% der Messstellen waren alle Anforderungen an ein tolerantes NZM-Design erfüllt (n = 806). Diese Messstellen liegen vorwiegend im Norddeutschen Tiefland und angebaute Kulturen wie Mais und Winterweizen sowie Grünland dominieren diese Standorte. Informationen wie bspw. die Grundwasserfließrichtung, Fließ- und Transportgeschwindigkeiten, sowie die Charakteristik der Böden im Anstrom müssen für eine abschließende Bewertung herangezogen werden. Steuernde Einflussfaktoren, die die intrinsische Vulnerabilität eines Standortes definieren, lassen sich aus den ML-Modellergebnissen nur bedingt herleiten für die NZM-Auswahl. Aus diesen Gründen ist für die Standortsuche zu empfehlen, die Mindestkriterien nach König et al. (2020) zu verwenden sowie auf direkt am Standort erhobene Daten und auf großmaßstäbliche Geodaten zurückzugreifen. Die behördliche lokale Expertise in den Bundesländern wird in diesem Zusammenhang ebenfalls empfohlen.

In Bezug auf die Methode SDLM (spatial distributed leaching modelling) empfehlen wir, die Vulnerabilität von NZM-Standorten in Deutschland nicht mit Modellen zu bewerten, die auf Geodaten mit großen Maßstabszahlen basieren. Standortbezogene Modellierungen mit GeoPELMO auf Basis von großmaßstäblichen Daten, bspw. Bodendaten der BK 1:5000 zu verwenden, erscheint uns zielführender hinsichtlich einer Einschätzung der Versickerungsbedingungen vor Ort. Modelle, die EU-weit mit Eingangsdaten im noch kleineren Maßstab (bspw. 1:1.000.000) arbeiten, sind daher mit Bezug auf die Vorhersagegüte einer standortbezogenen Vulnerabilität als kritisch zu bewerten. Darüber hinaus bedarf es weiterer Untersuchungen, erarbeitete Modelle bzgl. ihrer Treffsicherheit in Sachen Vulnerabilität zu überprüfen.

Summary

Background and objectives

For the authorization of plant protection products (PPPs), targeted monitoring studies are becoming increasingly important as highest tier studies in groundwater risk assessments. In context of an EU-wide harmonized risk assessment, the estimation of the vulnerability of monitoring sites and wells in relation to entries of active substances and their degradation products into groundwater are essential. One objective of the project was to evaluate, which factors significantly determine the vulnerability of groundwater in relation to entries of active substances and metabolites of PPPs in Germany. Based on this, the question of how the vulnerability of groundwater to PPP inputs can be quantified, classified and spatially predicted should be answered. In a further step, it should be examined whether the model results can be transferred to the development of a nationwide index for the vulnerability of groundwater in relation to PPPs. Linked to this is the question of new findings on relevant site characteristics that could influence the selection of monitoring sites for targeted post-registration monitoring studies (NZM) and what recommendations can be derived from this.

Database

In order to answer the questions, groundwater quality data obtained from the thirteen federal states for a large number of active substances and metabolites of PPPs were compiled, harmonized and evaluated in a comprehensive database in this project. The database was linked with information on substance and site properties from further data sources. In total, over 26,000 monitoring wells with over 500 substances are contained in the database. The oldest measurement was carried out in the year 1973, but the greatest measurement density and the broadest range of substances was recorded from 2010 onwards. At first, active substances were primarily in the spotlight but metabolites were increasingly included in the investigation program from 2007 onwards. The monitoring strategies, i.e. the number and type of substances searched for as well as measurement intervals and spatial coverage differ significantly depending on the federal state and over time. Most of the substances investigated are herbicides, followed by insecticides and fungicides and other substance classes. The metabolites investigated are mostly metabolites of herbicides.

Looking at the entire database, concentrations above the respective limit of quantification (LOQ) were measured at least once for 315 different PPP active substances and metabolites at a total of 15,301 monitoring sites in all thirteen federal states over the entire period. This means that active substances and metabolites were detected at 58.4% of the monitoring sites in the data set. At the same time, 96.6% of all measurements contained in the data set (regardless of the substance or monitoring well) are below the LOQ. Exceedances of $0.1 \mu\text{g L}^{-1}$ of a total of 156 active substances and relevant metabolites are found in the overall data set at 6,256 monitoring wells (19.9%) in thirteen federal states. In the period from 2010 onwards, 129 active substances and relevant metabolites exceeded the limit value at a total of 4,377 monitoring sites (23.9% of the monitoring sites sampled from 2010 onwards).

In Cooke et al. (2024), the proportion of measurements in the five concentration classes ($< \text{LOD}$, $< \text{LOQ}$, < 0.1 , > 0.1 and < 1 , and $> 1 \mu\text{g L}^{-1}$) was investigated for active substances and relevant metabolites or for non-relevant metabolites based on selected site factors and in two time periods. The distribution of the concentration classes was presented with regard to four aquifer cavity types (porous, fractured, fractured/porous and fractured/karst aquifers), three land use classes (arable land/grassland, forest and buildings) and the depth of the upper filter edge below the top of the terrain in five depth ranges (0-2, 2-10, 10-20, 20-50, 50-100 m). For the period 2010-2019, it can be summarized that higher proportions of higher concentration ranges

occur in all aquifers, but more frequently in pore aquifers. Furthermore, the results point to the influence of land use in the vicinity of the monitoring site (proportionally higher concentration classes in arable land) and the filter depth (proportionally higher concentration classes in filters that are closer to the surface). Exceedances of $0.1 \mu\text{g L}^{-1}$ only showed bivariate correlations with site factors with a coefficient > 0.3 for a few substances (mainly metabolites). The number of site factors that show correlations > 0.3 at all is also reduced. As expected, there are positive correlations with the associated and other crops, possibly due to crop rotation, as well as the average nitrate and potassium concentration in the groundwater.

Machine learning methods

Machine learning comprises statistical methods in which computer algorithms develop models from data that can be used for prediction, among other things. The relevant rules and correlations are not predefined, but are explicitly taken from the data. The learning process is referred to as training and involves an iterative adjustment of model parameters as well as the comparison of model predictions based on the initial data, in which the deviation between data and prediction is successively reduced. In this project, two different algorithms - random forest and neural networks - were applied in a classification. The target variable and the explanatory variables (so-called features) were defined and compiled beforehand. The weighting and analysis of the influence of the features on the prediction of the target variable were main objectives and were calculated by using appropriate methods.

The different and in some cases insufficient temporal resolution of the measurement data and explanatory variables demanded a spatial, temporally aggregated model approach was chosen. The strongly left-censored measurement data of the active substances and metabolites made it difficult to use the substance concentrations directly as target variables and had a decisive influence on the selection of substances for which models could be trained and calculated. The substance measurements were therefore divided into concentration classes and these were predicted at the measuring points for each substance (classification). The binary concentration classes are concentration class 1) $\leq 0.05 \mu\text{g L}^{-1}$ and a concentration class 2) $> 0.05 \mu\text{g L}^{-1}$. Sufficient data for robust models (sufficient values above the limit of quantification) were available for eight metabolites. Consequently, binary classification models were trained using two architectures for the metabolites desphenyl chloridazon, methyl-desphenyl chloridazon, metolachlor ESA, metolachlor OXA and NOA 413173 as well as metazachlor ESA, metazachlor acid and trifluoroacetic acid (TFA). These metabolites are predominantly relatively mobile and persistent degradation products of active substances with a long application history. The features used (reduced feature selection) were the area of the respective main or secondary crops in a 1-km radius around the monitoring well, the filter top edge, the average groundwater recharge, the sand content and the organic carbon content in the soil, the macropore class, hydraulic conductivity, and cavity type of the aquifer.

In a so-called extended feature selection, in addition to the above-mentioned features, the large hydrogeological area, long-term features for air temperature and for precipitation, terrain height, slope, rock type and geochemical type of the aquifer, several features for locating the measuring point in the surface catchment area, soil textural information of silt and clay content, the size of arable land and the long-term average percolation rates were also used. Furthermore, features derived from the groundwater chemistry measurements at the monitoring sites, which can be seen as additional indicators for groundwater quality, were also used (e.g., nitrate, ammonium, magnesium, sulfate, chloride, potassium and sodium). Due to the high proportion of left-censored measurements, the inclusion of active substances was only possible to a limited extent. Both feature selections were also calculated for individual active substances (bentazone,

atrazine and simazine) and for three groups of active substances summarized according to their mobility (K_{Foc} values) (very mobile and mobile; moderately mobile, less mobile).

Results of machine learning

Metabolites

The binary metabolite models of the reduced feature selection largely show model accuracies that are significantly higher than random assignments (Macro-F1 score > 0.7) and the majority of the monitoring sites (~80%) could be predicted correctly. Thus, with a manageable number of spatial and temporally aggregated features, spatial models for eight metabolites could be trained with good prediction quality.

Regardless of the algorithm, the respective main culture was identified as the most important feature for almost all metabolites. Other subsequent features include the filter top edge, sand content, groundwater recharge and, in some cases, secondary crops. The order of these features may differ slightly depending on the method. With TFA, the filter top edge and the sand content are classified as the most important, followed by numerous crops. The hydrogeological features and the macropore class were assigned a low feature weighting for all substances. Across all metabolites, there is a correlation that a larger area of the main culture is associated with the higher concentration class and a deeper filter top edge with the lower concentration class. For the features sand content and groundwater recharge, there are differences between the compounds, which can partly be linked to the distribution of these features in the respective cultivation areas. We assume that the macropore class - similar to the sand content, only reversed - indicates the distribution of soil properties in cultivation areas.

The addition of further features led to a reduction in the number of available monitoring sites for all compounds. The model grades were nevertheless higher than with the reduced feature selection. The respective main culture is still the most important feature for almost all substances. The average nitrate, sulfate and potassium contents in the groundwater are also very relevant. These features have an influence on the prediction of the concentration classes comparable to that of the crop and can probably also be interpreted as indicators for agricultural areas. The groundwater chemical features show an influence on the substance concentrations comparable to that of land use. For most metabolites, the nitrate concentration in groundwater is higher in cultivation areas, so it can be assumed that the groundwater chemical features are spatially related to the cultivation of the main culture. In combination with the analysis of the feature effects, it can be concluded that these groundwater chemical parameters can be seen as indicators of agricultural activity and possibly as an indication of the intensity of this activity. Although the presence of nitrate can also be an indication of increased infiltration potential, in this context it can be assumed that the models learn relationships primarily on the basis of spatial association. Other features (long-term meteorological conditions, altitude, etc.) can also refer to agricultural areas due to their interrelationship (multi-collinearity).

Active substances

The prediction qualities of the mobile and very mobile or less mobile active substance group models (reduced feature selection) are not robust enough with regard to the prediction of the higher concentration class, which is why the feature importance of these models is subject to uncertainties. However, the models of the mobile and very mobile active substances show a feature importance that is similar to that of the eight metabolites. The moderately mobile active substances assign a higher weighting to the hydrogeological and soil features than to the arable land feature. In the models of the extended feature selection, a slightly better model quality was

achieved for all substance groups and the individual substances. The nitrate content is the most important feature for the moderately mobile and less mobile active substances.

Classification of the results in the context of vulnerability and comparison with other vulnerability methods

The concentrations found in the groundwater samples represent several overlapping processes. Data on the actual application of PPP active substances are the most important link between land use and pesticide concentrations in groundwater, but were not available. The quantification of vulnerability (high, moderate, low) is difficult to decouple from the influence of actual application if no long-term data is available on the latter. Even undetectable or very low concentrations are therefore not necessarily synonymous with low intrinsic vulnerability, as the substance may not have been applied. Due to this limitation, there are limits to the assessment of groundwater vulnerability from the available analysis, particularly with regard to the development of a vulnerability index from those monitoring data.

Ultimately, it is concluded that no robust vulnerability index can be developed for PPPs on the basis of the available data. The next step should therefore be to examine whether existing vulnerability concepts could be evaluated using monitoring data. On the one hand, a comparison was made between the protective function of the groundwater cover (according to Hölting et al., 1995 and LAWA) and the monitoring data, and on the other hand between GeoPELMO DE simulation results and the monitoring data. The simulated leachate concentrations were divided into four spatial percentile classes.

The comparison between the protective function of the groundwater cover and the concentration classes of metabolites at the monitoring sites in the FARM dataset showed no correlation. Only in the case of active substances was there a tendency for sites with an “unfavorable” protective function to have proportionately more monitoring sites with exceedances. At the same time, a certain percentage of monitoring sites with exceedances occur at sites in all three categories of protective function. The “favorable” category includes sites where it is estimated that it can take more than 10, and in some cases more than 25 years for a substance to reach the groundwater. The metabolites considered are exclusively mobile and very mobile, some of which originate from active substances with a history of use spanning years and decades. It is assumed that the description of intrinsic vulnerability (according to Hölting et al. 1995 and the LAWA) is not sufficiently suitable for these metabolites with high mobility, which originate from existing substances. The degradation products have now reached numerous aquifers of different categories and occur ubiquitously.

The comparisons between monitoring data and the process-oriented model results with GeoPELMO DE largely showed no correspondence. Basically, it can be observed that for all active substances and metabolites considered, there are measuring points with findings in all spatial percentile classes. For metazachlor ESA, GeoPELMO DE (version FOCUS L; leaching, which means infiltration occurs due to matrix flow) shows the expected correlation (proportionally more overestimates in percentile class 75-100), while this is not clearly evident for GeoPELMO DE (version FOCUS L+R+M+D; beside from leaching the processes runoff, macropore flow and drainage are incorporated). For metazachlor acid, both model configurations show a tendency towards proportionally more overestimates in the 75-100 percentile class. However, there is no correlation for metolachlor ESA and metolachlor OXA. Among the active substances, there is an opposite correlation for bentazone with a higher proportion of measuring points with findings in the 0-25 percentile class in both model configurations. In the case of metolachlor, metazachlor and terbuthylazine, there is no correlation. A slight tendency towards an expected

correlation can at most be discussed for terbuthylazine for the GeoPELMO DE model configuration (FOCUS L+R+M+D).

Overall, it can be concluded that this comparison showed no correlation between the proportion of monitoring sites with findings and the higher spatial percentile class for most compounds (seven metabolites and four active substances). Monitoring sites with exceedances of the selected concentrations were found in all spatial percentile classes. An advantage of GeoPELMO DE (FOCUS L+R+M+D) with additional processes like runoff, macropore flow and drainage over the FOCUS-based version of GeoPELMO DE (FOCUS L) cannot be derived from this. Interpreting this result should take into account that the developers of GeoPELMO DE postulate a limited influence of the additional processes runoff, macropore flow and drainage on the model-based prediction of leachate concentrations for metabolites. This could be the reason for the comparatively small differences in our comparisons of the monitoring data with both model versions for the metabolites investigated.

In summary, it can be said that the considerations regarding the classification of the ML model results with regard to their significance in the vulnerability assessment and the comparison of the index and process-based methods with the monitoring data raise numerous questions. In the case of the index method under consideration, it is assumed that this is inadequate for substances with a long history of use and high mobility. The discrepancy between the simulations from spatial distributed leaching models and the monitoring data primarily highlights the question of scale: Are the simulated leachate water concentrations in the soil and the concentrations in the groundwater comparable in terms of time and space? For these reasons, no clear recommendations can be derived from this project as to which of the two available methods is more suitable, e.g. to identify vulnerable sites for targeted post-registration studies.

The ML models for mobile and very mobile metabolites or active substances tend to identify the agricultural area (specific crop or arable land in general) as one of the most important, if not the most important feature. Hydrogeological properties play a clearly subordinate role. This observation is consistent with the deviation from the conservation function categories.

This was not always the case for the active ingredient and active ingredient group models. Especially for the moderately mobile active substances, other features (soil, hydrogeology) were ranked higher. The comparison between Hölting and the findings of the active substances tended to show that the protective function of the groundwater cover can better describe active substance findings, or that the active substances may be better suited for the spatial differentiation of intrinsic vulnerability than the metabolites investigated with their properties. The specific role played by the mobility of the compounds could not be conclusively clarified. This is underlined by the similar feature weighting between metabolites and mobile active substances. The mobile and very mobile active substance models also showed a low feature weighting of the hydrogeological features, while the feature arable land was classified as relatively important. Caution is required when comparing active substances and metabolites, as the latter are only formed in the soil and may be stored there for a longer period of time. We suspect that it may be possible to learn more about intrinsic vulnerability from less mobile active substances. At the same time, active substances occur in lower concentrations (high proportion of left-censored data) that no meaningful comparisons can be made or robust models can be trained. A lowering of the LOQ along with a nationwide harmonization of LOQs would enable a more streamlined analysis to train ML models with PPPs.

Recommendations for targeted post-registration monitoring studies (NZM)

The results of the harmonized database enable to preselect potential monitoring sites for targeted post-registration monitoring studies (NZM) in Germany. Upon the 26,000 monitoring wells, which were analyzed within the FARM project, 7,675 do not meet the criteria for NZM-compatibility. Reasons for exclusion are sites with deep water table depths and/or top filter edges away from the groundwater surface. Data about the water table (n = 6,377) were predominantly missing compared with data about the monitoring well (n = 1,407) and excluded evaluation of these sites. Both information were missing in 6,686 cases. In only 3% of all cases all tolerant criteria were met for an NZM-design (n = 806). These sites are located in Northern Germany and main cultivated crops are maize and winter wheat as well as grassland. Information about the groundwater flow direction, hydraulic conductivities, and soil characteristics located in lateral flow direction need to be assembled for a coherent verification of the NZM-design. Driving features, which determine the intrinsic vulnerability, could not reliably transferred from the ML results to be supportive for NZM judgement. Therefore, it is recommended that the minimum criteria according to König et al. (2020) should be used for targeted site selection. Furthermore, site assessments should consider data collected directly at the site with geodata at better spatial resolution compared with nationwide available data. Expertise from local authorities in the federal states is also recommended in this context.

With regard to SDLM (spatial distributed leaching models), we recommend that the vulnerability of targeted monitoring sites for post-registration studies in Germany should not be assessed using models based on small scale geodata. Site-specific modelling with GeoPELMO based on large-scale data, e.g. soil data from the BK 1:5000, seem to be more suitable in terms of assessing site-specific leaching conditions. EU-wide models based on input data on a lower spatial resolution than national level (e.g. 1:1,000,000) must therefore be assessed as critical with regard to their predictive quality of site-specific vulnerability. In addition, further investigations are needed to check the accuracy of developed models with regard to vulnerability.

1 Hintergrund und Zielsetzung

Gesetzlicher Rahmen

Der Schutz der natürlichen Ressource Grundwasser ist für die Wasserversorgung und den Erhalt von Ökosystemen unabdingbar. In Deutschland wird Trinkwasser zu rund 70% aus dem Grundwasser gewonnen. Die EU-Wasserrahmenrichtlinie (EU 2000) und die EU-Grundwasserrichtlinie (EU 2006) regeln Kriterien für den guten chemischen und mengenmäßigen Zustand des Grundwassers. Die Grundwasserverordnung (GrwV 2022) spiegelt die Umsetzung in nationales Recht wider. Wenn Wasserqualitätsstandards nicht eingehalten werden, liegt dies u.a. an Verunreinigungen mit Pflanzenschutzmitteln (PSM) und ihren Abbauprodukten. In der EU erreichten 6,5% der Grundwasserkörper keinen guten Zustand, vorwiegend aufgrund von Grenzwertüberschreitungen durch Nitrat und PSM (Mohaupt et al. 2020). Auch wurde eine hohe Belastung von kleinen Fließgewässern mit PSM in landwirtschaftlich geprägten Gebieten festgestellt (Liess et al. 2019). Der Schutz des Grundwassers hat im Zulassungsverfahren für PSM einen besonders hohen Stellenwert. PSM dürfen nach Artikel 4 der Pflanzenschutzmittelverordnung (EG) Nr. 1107/2009 keine schädlichen Auswirkungen auf das Grundwasser haben. Daraus ergibt sich die Notwendigkeit einer Risikobewertung der PSM-Konzentrationen im Grundwasser sowie die Durchführung von Maßnahmen zur Reduktion von Einträgen. Eine prospektive Bewertung wird in der Regel anhand von Modellierungen vorgenommen, aber auch Monitoringstudien finden Eingang in die Bewertung. Gemäß der EU-Pflanzenschutzmittelverordnung (1107/2009) und der Trinkwasserrichtlinie (EU-Trinkwasserrichtlinie 2020/2184) gilt für PSM-Wirkstoffe ein Grenzwert von $0,1 \mu\text{g L}^{-1}$, der sich auch in der EU-Grundwasserrichtlinie 2006/118/EC wiederfindet.

Wirkstoffe und Metaboliten von Pflanzenschutzmitteln

Wirkstoffe werden zu Abbauprodukten (Metaboliten) zersetzt. Diese werden im Pflanzenschutzrecht nach europäischen Richtlinien (SANCO/221/2000; EC, 2021a) als relevante Metaboliten (rM) bzw. als nicht-relevante Metaboliten (nrM) bewertet. Relevante Abbauprodukte haben eine mit ihrer Ausgangssubstanz/Muttersubstanz vergleichbare pestizide Wirkung und/oder besitzen Eigenschaften, die eine toxikologische Gefährdung von Verbrauchern und Gewässerorganismen bedingen können. Die Einschätzung von Metaboliten als relevant oder nicht-relevant wird von der Europäischen Lebensmittelbehörde (EFSA) in den EFSA Conclusions zur (Wieder)Genehmigung von PSM-Wirkstoffen in der EU veröffentlicht. Die Europäische Chemikalienagentur (ECHA) entscheidet in diesem Zusammenhang letztlich über die Legaleinstufung von Wirkstoffen anhand ihrer toxikologischen Eigenschaften. Grundsätzlich wird bei fehlender oder unzureichender Datenlage für einzelne Metaboliten die gleiche toxische Gefährdung wie für die Ausgangssubstanz/Muttersubstanz angenommen, diese kann anschließend durch Studien der Antragssteller entkräftet werden. Für Abbauprodukte, für die eine Toxizität nicht ausgeschlossen werden kann und für die noch keine abschließende Bewertung vorliegt, steht eine eindeutige Zuordnung in rM oder nrM noch aus. Dafür wurde national nach (Banning et al. 2022) die Kategorie „xM“ eingeführt. Für relevante Metaboliten (rM) gilt der Grenzwert für Wirkstoffe von $0,1 \mu\text{g L}^{-1}$. In Summe sollen rM und Wirkstoffe einen Wert von $0,5 \mu\text{g L}^{-1}$ im Grundwasser nicht überschreiten. Nicht relevante Metabolite sind jedoch nicht zwangsläufig als unbedenklich für die Umwelt und menschliche Gesundheit zu bewerten. Sie können mobil und persistent sein, d.h. leicht transportiert und langsam abgebaut werden (Adlunger et al. 2022). Zu vielen nicht-relevanten Metaboliten (nrM) liegen nicht genügend Daten vor und neue Studienergebnisse können zu einer Einstufung als rM führen. In den EU-Staaten werden nrM z.T. unterschiedlich reguliert. Laut der aktuellen EU-Trinkwasserrichtlinie

sind nationale Grenzwerte für nrM, ab denen Maßnahmen notwendig werden, durch die Mitgliedstaaten zukünftig festzulegen. In Deutschland wird für nrM der aktuell nicht rechtsverbindliche Vorsorgemaßnahmenwert von $10 \mu\text{g L}^{-1}$ und für einige Metaboliten im Trinkwasserrecht ein spezifischer Gesundheitlicher Orientierungswert (GOW) von 1 oder $3 \mu\text{g L}^{-1}$ angewandt.

Grundwasservulnerabilität

Die Vulnerabilität des Standortes hinsichtlich des Eintrags von PSM ist besonders maßgeblich, vor allem, wenn Monitoringergebnisse regulatorisch im Pflanzenschutzrecht verwendet werden. Grundwasservulnerabilität basiert zum einen auf der intrinsischen, d.h. der stoffunabhängigen, an die Eigenschaften der Bodenzone, der Überdeckung und des Grundwasserleiters gebundenen Vulnerabilität. Hinzu kommt die spezifische Vulnerabilität, die maßgeblich durch die Eigenschaften des Stoffes, wie Adsorptionsverhalten und Abbauraten, aber auch durch Ausbringmenge und -zeitpunkt beeinflusst wird. Im Grundwasserschutz werden Grundwasservulnerabilitätskarten verwendet, für deren Erstellung es unterschiedliche Methoden gibt. Dazu gehören u.a. Punktebewertungs- und Indexverfahren, wie bspw. DRASTIC (Aller et al. 1987); näheres dazu in Abschnitt 5. Allerdings wird in dieser Methodengruppe fast ausschließlich die intrinsische Vulnerabilität eines Aquifers bewertet. Grundsätzlich besteht Forschungsbedarf bei der Entwicklung von Bewertungsmethoden, welche die Kombination aus intrinsischer und spezifischer Vulnerabilität, insbesondere im Kontext des PSM-Eintrags und der landwirtschaftlichen Nutzung, betrachtet (Gimsing et al. 2019).

Die Verwendung unterschiedlicher Ansätze für eine Vulnerabilitätsbewertung von Monitoringmessstellen und -standorten im Kontext Pflanzenschutz werden auch auf EU-Ebene intensiv diskutiert (Gimsing et al. 2019; Tiktak et al. 2020). Dabei werden räumliche Modellansätze auf Basis der FOCUS-Modelle PEARL, PELMO und MACRO für eine Vulnerabilitätsbewertung in Betracht gezogen, die das Abbau- und Adsorptionsverhalten von PSM im Boden in Abhängigkeit von Anwendungsmenge und -zeitraum prozessorientiert berücksichtigen und eine erwartbare Versickerungsneigung berechnen. Der Vorteil dieser vorgeschlagenen Ansätze ist, dass sie stoffspezifische Eigenschaften sensitiv berücksichtigen und damit der Bewertung einer spezifischen Standortvulnerabilität gegenüber PSM generell zuträglich sind. Darin unterscheiden sie sich von einfachen Indexansätzen, die vorwiegend eine intrinsische Vulnerabilität abbilden. Alle diese für Vulnerabilitätsanalysen in Betracht gezogenen PSM-Modelle bilden jedoch überwiegend die Prozesse in der ungesättigten Bodenzone ab und berücksichtigen kaum den Grundwasserflurabstand bzw. den gesamten Aufbau und die gesamte Mächtigkeit der ungesättigten Zone und die damit zusammenhängenden Versickerungs- bzw. Verweilzeiten. Hydrogeologische Gegebenheiten der gesättigten Grundwasserzone werden gänzlich unberücksichtigt gelassen. Ein weiteres Defizit ist, dass die existierenden räumlichen Modellansätze bisher kaum anhand größerer Monitoringdatensätze validiert wurden.

In der Pflanzenschutzmittelverordnung (EG) 1107/2009 ist darüber hinaus verankert, dass die Prüfung von PSM in der EU nach weitestgehend einheitlichen Bewertungsmaßstäben unter gleichzeitiger Berücksichtigung nationaler Besonderheiten erfolgen soll. Mit der Zustimmung der Mitgliedstaaten im Oktober 2014 im Ständigen Ausschuss der Kommission (KOM) im Rahmen des FOCUS-Stufenkonzept (FOCUS 2009/2014) der Grundsatz harmonisierter Bewertungsmaßstäbe vor allem für die modell- und szenarienbasierte Risikobewertung des Grundwassers weiter umgesetzt. Eine harmonisierte Leitlinie zur Bewertung von PSM-Monitoringdaten für die Risikobewertung des Grundwassers in der EU existiert derzeit nicht (EFSA PPR Panel et al. 2023), wobei eine Bewertung anhand von Monitoringdaten in den einheitlichen Bewertungsmaßstäben der Regulation 546/2011 (Anhang C 2.5.1.2) durchaus eingeräumt wird. Der Bedarf einheitlicher Bewertungsansätze, die über eine Berücksichtigung

nationaler Besonderheiten hinausgehen, ergibt sich durch zielgerichtete Monitoringstudien und Auswertungen von Monitoringdaten, die von den Antragstellern zunehmend für die Wiedergenehmigung von PSM-Wirkstoffen auf EU-Ebene und in zonalen bzw. nationalen Dossiers zur Zulassung von PSM-Anwendungen eingereicht werden. Die fehlenden übergeordneten Bewertungsmaßstäbe für Monitoringdaten im Pflanzenschutz in der EU sowie die nationale bzw. föderale Verantwortung für Grundwassermonitoringprogramme begründen die Notwendigkeit, fachlich grundsätzliche Fragen zur Grundwasservulnerabilität gegenüber PSM auf nationaler Ebene zu untersuchen.

Risikobewertung im PSM-Vollzug

Im Rahmen der nationalen Grundwasserrisikobewertung werden für einzelne PSM bereits Monitoringdaten bei Entscheidungen über die Zulassungsfähigkeit in Verbindung mit Managementauflagen berücksichtigt. Fundaufklärung (FA) und Nachzulassungsmonitoring (NZM) sind die beiden seit 2002 etablierten und in den letzten Jahren überarbeiteten Verfahren in Deutschland, in denen Bund und Länder eng zusammenarbeiten (Aden et al. 2002; Fundaufklärungsverfahren 2016; König et al. 2020). Im Fundaufklärungsverfahren werden auffällig gewordene Messstellen der Länder und der Wasserversorger einer retrospektiven Ursachenanalyse für erhöhte Wirkstoff- bzw. Metabolitenkonzentrationen unterzogen. Beim NZM werden für die zu untersuchenden PSM zielgerichtete, langjährige Studien, mit dokumentierten PSM-Anwendungen an wenigen, gut untersuchten Grundwassermessstellen (GWM) mit hoher Vulnerabilität durchgeführt. Im derzeit etablierten Verfahren werden vorrangig GWM aus Ländermessnetzen in NZM-Studien verwendet. Beide Studientypen dienen in der nationalen Regulierung von PSM-Anwendungen dazu, die modellgestützte Risikobewertung in der Realität zu überprüfen, ggfs. adverse Effekte zu identifizieren und den Grundwasserschutz im Pflanzenschutz zu verbessern. Trotz der Neugestaltung beider Verfahren (FA und NZM) sind der Interpretation solcher Monitoringdaten für regulatorische Entscheidungen bisher noch Grenzen gesetzt. Ein Grund dafür ist, dass entsprechende Daten im Zulassungsverfahren von PSM immer in einem bundesweiten Kontext zu bewerten sind, obwohl die Daten häufig nur für eine begrenzte Anzahl an Messstellen aus einzelnen Bundesländern vorliegen. Darin liegt die Notwendigkeit begründet, die Repräsentativität und die Vulnerabilität von Messstellen und Monitoringstandorten bezüglich des bundesweiten Versickerungsrisikos von Wirkstoffen und Metaboliten fachlich gut zu beschreiben. Für diesen Zweck bedarf es der Untersuchung von Kriterien, welche die Vulnerabilität des oberflächennahen Grundwassers in landwirtschaftlichen Gebieten gegenüber PSM-Einträgen maßgeblich bestimmen.

NZM-Studien dienen der Risikoüberprüfung, inwieweit konkrete PSM-Anwendungen zu Einträgen in den obersten Grundwasserleiter führen. Dieser Zweck unterscheidet sich von den Zielen des Wasserrahmenrichtlinie (WRRL)-Monitorings der Länder und bestimmt gleichzeitig die hohen Anforderungen an die Messstellenauswahl, an Informationen zum Anstromgebiet und an das Studiendesign solcher Studien. Nach den derzeit festgelegten Kriterien sind für NZM-Studien Gebiete mit Verweilzeiten von maximal 3 Jahren in der ungesättigten Zone und mit schnellen Transportraten in der gesättigten Zone vorrangig von Interesse.

Hinsichtlich der Böden und Deckschichten erfüllen Lockergesteinsgebiete mit durchlässigen Substraten ohne stauende Schichten die Voraussetzungen für eine Eignung im NZM-Verfahren, da hier ein entsprechend hoher Matrixfluss und Stofftransport in der ungesättigten Zone gewährleistet ist. Vor allem in sandigen Böden und Lockergesteinsgebieten ist ein hohes Versickerungspotential von PSM zu erwarten und gut zu erfassen. Doch auch Gebiete mit bindigen, schluffigen und tonigen Böden bzw. Substraten, die Makroporenbildung und präferentielles Fließen (Makroporenfluss) begünstigen, können für Grundwassereinträge von PSM anfällig sein (Rosenbom et al. 2015) und sollten im NZM miterfasst werden. Hier bestehen

jedoch starke Unsicherheiten in der Anwendbarkeit des für NZM-Studien vorgesehene Studiendesigns bezüglich einer sinnvollen Messstellenauswahl und -anordnung. Die Auswahl von Probenahmepunkten für NZM-Studien in freien Porengrundwasserleitern birgt vermutlich geringere Unsicherheiten bei der Interpretation der Ergebnisse gegenüber Untersuchungsgebieten in denen Karst auftritt (Gimsing et al. 2019). Kluft- oder Karstaquifere kommen in Deutschland jedoch ebenfalls häufig vor und können ebenfalls als vulnerabel gegenüber PSM-Einträgen gelten. Hier fehlen bisher Vorstellungen, welche der bundesweit vorkommenden Festgesteinsgebiete besonders vulnerabel gegenüber Pflanzenschutzmitteleinträgen sind und mit welchem Studiendesign auch diese Gebiete in NZM-Studien zukünftig berücksichtigt werden können. Diese offenen Fragen im NZM-Verfahren begründen die Notwendigkeit, die Vulnerabilität des Grundwassers gegenüber PSM in diesem Projekt eingehender zu untersuchen.

Projektziele

Grundwassermonitoringdaten spielen eine zentrale und zunehmende Rolle bei der Risikobewertung des Grundwassers im Rahmen der Zulassung von einzelnen PSM. Einschätzungen zur Vulnerabilität von Standorten und Messtellen gegenüber Einträgen von PSM-Wirkstoffen und ihren Abbauprodukten ins Grundwasser sind dafür essentiell. Ziel des Projekts ist es, eine Strategie zu entwickeln, um die regulatorischen Anforderungen der PSM-Zulassung mit vorhandenen und zukünftigen Monitoring-Programmen besser in Einklang zu bringen. Dafür wurden bundesweit harmonisierte Monitoringdaten mit maschinellen Lernverfahren analysiert, um Standortfaktoren der Grundwasservulnerabilität zu identifizieren und deren Einfluss zu quantifizieren. Im Projekt wurde die Entwicklung eines statistischen Vulnerabilitätsindex, der über bestehende Ansätze hinausgeht und die Limitationen von Versickerungsmodellen ausgleicht, geprüft. Der Index soll intrinsische, stoffspezifische und kulturspezifische Aspekte der Vulnerabilität berücksichtigen und zur Ausweisung geeigneter Monitoringgebiete beitragen. Die Ergebnisse unterstützen die fachlich fundierte Auswahl von Messstellen und liefern Erkenntnisse für die Optimierung zukünftiger Studien im Zulassungsverfahren.

Das Projekt war in vier Arbeitspakete unterteilt:

Arbeitspaket 1: Erfassung PSM-Eintragsituation

- ▶ Zusammenstellung existierender Monitoringberichte
- ▶ Datenabfrage bei den Bundesländern und Aufbau Datenbank PSM-Messstellen
- ▶ Ermittlung von Fruchtfolgen zur Abschätzung potentieller historischer PSM-Einträge
- ▶ Statistische Auswertung PSM-Datenbank und Simulation von Erwartungskonzentrationen mit GeoPELMO DE

Die Schritte der Zusammenstellung der Datenbank sowie eine Übersicht aller vorhandener Parameter wird in Kapitel 2.1 beschrieben, die Auswertung der Datenbank erfolgt in Kapitel 3.1. Der Abgleich zwischen GeoPELMO-Simulationen und Monitoringdaten wird in Kapitel 5.2 gezeigt. Der Abgleich mit existierenden Monitoringberichten fand u.a. in Cooke et al. (2024) statt.

Arbeitspaket 2: Kriterien für PSM-Vulnerabilität

- ▶ Identifizierung und Gruppierung der für die Vulnerabilität entscheidenden Faktoren

► Bewertung Einfluss Klimawandel

Die Identifikation der relevanten Standortfaktoren mittels maschinellen Lernens wird in Kapitel 2.2.3 methodisch erklärt. Die Ergebnisse werden in 3.2 gezeigt und in Kapitel 4 kritisch diskutiert. Der Einfluss des Klimawandels wird kurz in Anhang A.1 behandelt.

Arbeitspaket 3: Entwicklung und Validierung PSM-Indexverfahren zur Vulnerabilitätsbewertung

- Kompilation und Bewertung existierender Indexverfahren und prozessorientierter Modellansätze
- Entwicklung PSM-Indexverfahren auf Bundesebene
- Validierung und Kalibrierung des PSM-Indexverfahrens

Arbeitspaket 4: Ableitung von Empfehlungen

- Empfehlungen für die bundesweite Anwendung eines PSM-Indexverfahrens zur Bestimmung der Vulnerabilität von Monitoringstandorten
- Empfehlungen zur Verwendung bzw. Weiterentwicklung von GeoPELMO DE zur Vulnerabilitätsabschätzung von Standorten
- Empfehlungen für NZM-Studien im Rahmen des PSM-Vollzugs

Existierende Indexverfahren werden in Kapitel 5.1 und 5.2 behandelt. Eine Bewertung der Ergebnisse hinsichtlich einer Entwicklung von PSM-Indexverfahren wird in Kapitel 5.3 diskutiert, dies umfasst auch die in Arbeitspaket 4 geforderten Empfehlungen.

2 Methoden

2.1 Datengrundlage und Datenaufbereitung

In diesem Projekt wurden erstmalig Monitoringdaten der Länder für eine Großzahl an Stoffen, nach denen an einigen tausend Messstellen gesucht wird, in einer Datenbank zusammengeführt, harmonisiert und ausgewertet. Diese wurden mit Informationen zu den Stoff- und Standorteigenschaften aus anderen Datenquellen verschnitten und aufbereitet. Ein beträchtlicher Anteil der in diesem Projekt geleisteten Arbeit floss in die Datenharmonisierung (BGR et al. 2025). Im Folgenden wird die Datengrundlage beginnend bei den von den Ländern bereitgestellten Daten vorgestellt.

2.1.1 Wirkstoff- und Metabolitenkonzentrationen sowie Stammdaten der Grundwassermessstellen

Die zuständigen Staatlichen Geologischen Dienste und Landesumweltämter der 13 Flächenbundesländer wurden gebeten, Messungen von Wirkstoff- bzw. Metabolitenkonzentrationen an Grundwassermessstellen, Brunnen und Quellen bereitzustellen. Aus einem Teil der Bundesländer konnten zusätzlich zu den Landesmessstellen auch Daten von Wasserversorgungsunternehmen (WVU) geliefert werden.

Die ursprünglich bei den Bundesländern angefragten Stammdaten sind:

- ▶ Messstellen-ID
- ▶ Messstellen-Bezeichnung
- ▶ Gemeinde
- ▶ Koordinaten
- ▶ Geländeoberkante (GOK) [m NN]
- ▶ Messpunkthöhe [m NN]
- ▶ Endteufe [m u. Ansatzpunkt]
- ▶ Erfassungsdatum
- ▶ Filteroberkante unter GOK [m]
- ▶ Filterunterkante unter GOK [m]
- ▶ Ausbautyp/Messstellenart
- ▶ Rohr-/Filtermaterial
- ▶ Kennzeichnung Messnetz (Landesmessnetz, Wasserversorger etc.)

Auf die Messdaten der Wirkstoff- und Metabolitenkonzentrationen und Besonderheiten bei ihrer Auswertung wird gesondert in Kapitel 2.2.1 eingegangen. Die Auswertung des in der Datenbank vorhandenen Stoffspektrums erfolgt in 3.1.2.

2.1.2 PSM-Stoffeigenschaften

Physikochemische Stoffeigenschaften können herangezogen werden, um den Transport, das Sorptionsverhalten und den Abbau der Stoffe zu beschreiben. Prozesse wie Hydrolyse, Photolyse, Löslichkeit in Wasser, Abbauprodukte in Wasser und Sediment stammen aus Laborstudien und werden unter anderem in der *Pesticide Property Data Base* (PPDB) gesammelt (Lewis et al. 2016). Mobilität und Transport werden durch den Adsorptionskoeffizienten (K_{Foc}) im Boden beschrieben. Die Abbauprodukte im Boden (DT_{50}) gilt als Maß für die Persistenz einer Substanz. In der PPDB fehlende K_{Foc} - und DT_{50} -Werte wurden mit Werten aus aktuellen EFSA-Conclusions (persönliche Mitteilung, UBA, Stand 2022) aufgefüllt. Zur besseren Handhabung und systematischen Einordnung der Stoffe nach ihrem unterschiedlichen Transport- und Abbauverhalten, wurden die Wertebereiche des K_{Foc} und der DT_{50} in jeweils vier Klassen eingeteilt (Tabelle 1) (persönliche Mitteilung, UBA, 2022). Die Aufteilung der Stoffe in die insgesamt 16 möglichen Kombinationen aus Persistenz- und Mobilitätsklassen wird in Abbildung 18 dargestellt.

Tabelle 1: Übersicht der Persistenz- und Mobilitätsklassen anhand des Adsorptionskoeffizienten (K_{Foc}) und der Abbauprodukte im Boden (DT_{50}).

dt_{50} im Boden [Tage]	Persistenzklasse	K_{Foc} [$mg L^{-1}$]	Mobilitätsklasse
> 365	sehr persistent	< 30	sehr mobil
> 100	persistent	> 30, < 100	mobil
> 30	moderat persistent	> 100, < 500	moderat mobil
< 30	nicht persistent	> 500	wenig mobil

Der Adsorptionskoeffizient (K_{Foc}) und die Abbauprodukte im Boden (DT_{50}) wurden jeweils in vier Bereiche eingeteilt. Eine DT_{50} von mehr als 365 Tagen wurde als „sehr persistent“ klassifiziert, bei mehr als 100 Tagen als „persistent“, bei mehr als 30 Tagen als „moderat persistent“ und bei weniger als 30 Tagen als „nicht persistent“. Die Mobilitätsklasse „sehr mobil“ wurde bei einem K_{Foc} von unter $30 mg L^{-1}$ vergeben, zwischen 30 und $100 mg L^{-1}$ wird von „mobil“ gesprochen, zwischen 100 und $500 mg L^{-1}$ von „moderat mobil“ und ab $500 mg L^{-1}$ von „wenig mobil“.

2.1.3 Grundwasserchemische Messungen

Weitere grundwasserchemische Parameter, die in der harmonisierten Datenbank bereitgestellt wurden, sind die folgenden:

Aluminium, Ammonium, Arsen, Bor, Kalzium, Chlorid, gelöster organischer Kohlenstoff (DOC), Eisen, elektrische Leitfähigkeit bei $25^{\circ}C$, Fluorid, Gesamthärte, Hydrogencarbonat, Kalium, Kupfer, Magnesium, Mangan, Natrium, Nitrat-Stickstoff, Nitrat, pH-Wert, Phosphor (gesamt), Orthophosphat-Phosphor, Redoxpotential, Sauerstoff, Selen, Silizium, Sulfat, Wassertemperatur, Zink.

Je nach Datenverfügbarkeit konnte nur ein Teil dieser Parameter bundesweit ausgewertet werden.

2.1.4 PSM-Zulassung und Verkaufszahlen

Das Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) koordiniert Zulassung und Risikobewertung mittels der Zusammenarbeit der Fachbehörden Umweltbundesamt (UBA), Bundesinstitut für Risikobewertung (BfR) und Julius-Kühn-Institut (JKI). Das BVL hält umfangreiche Datenbanken zur Zulassungssituation von Wirkstoffen sowie bundesweite jährliche Verkaufszahlen bereit. Die Wirkstoffe der BVL-Datenbanken wurden mit den Stoffen

der Projekt-Datenbank des Forschungsvorhabens FARM abgeglichen, sodass Zulassungsinformationen (bis zu monatlicher Auflösung) sowie jährliche Verkaufszahlen in den Datensatz integriert werden konnten.

2.1.5 Standortfaktoren

Um eine große Anzahl an Messstellen bearbeiten zu können, wurde als erste Annäherung an ein Grundwassereinzugsgebiet ein kreisförmiger „Puffer“ von 1 km-Radius um die Koordinaten der Grundwassermessstelle gewählt. Eine Einzelbetrachtung jeder Messstelle und die Ableitung des Anstromgebietes wurde als zu umfangreich und nicht zielführend eingestuft. Trotz einer möglichen Über- oder Unterschätzung des Grundwassereinzugsgebiets aufgrund einer starren Geometrie hat sich ein Radius als sinnvoll erwiesen (Knoll et al. 2019; Manamsa et al. 2016). Je nach Datensatz, wurden Flächendaten z.T. auch direkt an der Koordinate extrahiert.

2.1.5.1 Hydrogeologie

Zu Beginn des Projektes wurden die Schichtverzeichnisse der Bohrungen an den Messstellen abgefragt. Die Rückmeldungen zeigten ein heterogenes Bild bezgl. der Datenverfügbarkeit und Kompatibilität. Schichtverzeichnisse werden u.a. in SEP2, SEP3 oder landesspezifischen Formaten vorgehalten; z.T. liegen sie als Kopien der Bohrprofilpläne und nicht in digitaler Form vor. Der Aufwand für die Harmonisierung der Schichtdaten wurde im Rahmen des Projektes für unverhältnismäßig befunden. Stattdessen wurde sich der Hydrogeologischen Übersichtskarte (BGR und SGD 2019) bedient. Diese umfasst u.a. Informationen zu Hohlraumart, Gesteinsart, Geochemie, Verfestigung und den hydraulischen Leitfähigkeitsklassen des obersten Grundwasserleiters. Vorteilhaft ist hierbei die bundesweite Vereinheitlichung der Aquifer-Eigenschaften und somit die Vergleichbarkeit der Daten, wenngleich die Auflösung von 1:250.000 zu einer Ungenauigkeit an der einzelnen Messstelle führen kann. Pro Messstelle wurde innerhalb des 1 km-Umkreises die jeweils flächenhaft größte Einheit ausgewählt. Zusätzlich wurde die Zugehörigkeit einer Messstelle zum jeweiligen hydrogeologischen Großraum (BGR und SGD 2015) extrahiert.

2.1.5.2 Wasserhaushalt

Die mittlere jährliche Sickerwasserraten wurden im 1 km-Radius um die Messstelle gemittelt (BGR 2023), genauso wie die mittlere jährliche Grundwasserneubildung (BGR 2019).

2.1.5.3 Grundwasserflurabstand

Die Bundesländer wurden gebeten, Grundwasserstandsdaten zu den PSM-Messstellen bereitzustellen. Aus den gelieferten Daten konnten für 8.254 Messstellen minimale, maximale und mittlere langjährige Grundwasserflurabstände zusammengestellt werden. Die Grundwasserflurabstände wurden danach gefiltert, ob eine Mindestanzahl von fünf Grundwasserstandsmessungen als Grundlage bereitstand. Der Anteil der Messstellen, für die Grundwasserflurabstände vorhanden sind, variiert in den Bundesländern. Die mittleren Grundwasserflurabstände sind von stark variabler Datenqualität, da sich die Zeitreihen in ihrer Länge, ihrem Start- und Enddatum und der Messfrequenz stark unterscheiden. Aus diesem Grund wurde der Grundwasserflurabstand nicht als Feature in den Machine Learning (ML)-Modellen verwendet. Der Grundwasserstand fließt z.T. in die Schätzung des Einflusses von Makroporen im Boden ein (2.1.5.4). Die Auswahl von potenziellen Messstellen für das Nachzulassungsmonitoring (NZM) (6) ist ebenfalls von der Datenqualität der Grundwasserflurabstände beeinflusst.

2.1.5.4 Boden

Bodenvariablen wurden aus der Bodenübersichtskarte im Maßstab 1:250.000 abgeleitet (BGR 2024). Sand-, Schluff- und Tongehalte der einzelnen Horizonte wurden gemäß der Mächtigkeit der Horizonte gewichtet und berechnet. Humusgehalte der Bodenhorizonte wurden in den Anteil an organischen Kohlenstoff (C_{org}) umgerechnet durch Division mit dem Faktor 1,72 und bei Torfen mit dem Faktor 2 (Ad-hoc-Arbeitsgruppe Boden 2005). Analog dazu wurde für die Lagerungsdichte jeweils die Mitte der Klasse verwendet (Ad-hoc-Arbeitsgruppe Boden 2005). Alle Parameter wurden tiefengewichtet für eine Profiltiefe bis 1 m unter GOK zusammengefasst. Diese umfassen den Sand-, Schluff-, und Tongehalt, pH-Wert, Humusgehalt und Boden- k_f -Wert und den C_{org} Gehalt. Weiterhin stehen die (nutzbare) Feldkapazität, potenzielle Wurzeltiefe sowie Grob- und Feinbodenvolumenprozent zur Verfügung. Makroporenklassen wurden anhand der Bodenarten zugeordnet, wobei eine Anpassung der Makroporenklassen mittels Grundwasserflurabstand mangels ausreichender und verlässlicher Daten im Projekt nicht umgesetzt wurde (Klein et al. 2019; Klein et al. 2025). Alle Flächen- und bis 1 m unter Bodenoberkante auch tiefengewichtete Bodeninformationen wurden mit dem 1-km Radius (Buffer) einer jeden Messstelle des Datensatzes verschnitten.

2.1.5.5 Klima

Meteorologische Daten wurden dem HYRAS Rasterdatensatz (DWD, 1-km-Auflösung, Wert extrahiert an Messstellenkoordinaten) für die Zeitspanne 1990-2020 entnommen. Niederschlagssummen, Lufttemperatur und Dürreindex wurden jeweils als monatliche, halbjährliche, jährliche, dekadische sowie 30-jährige Mittel-, Minimal- und Maximalwerte aggregiert.

2.1.5.6 Topographie

Topographische Indizes wurden aus Amatulli et al. (2018) (1 km Auflösung) entnommen. Die jeweiligen Werte wurden an Messstellen-Koordinaten extrahiert. Hierzu gehören u.a. die Geländehöhe, Hangneigung, Exposition, Geländerauigkeit, Profilkrümmung und weitere Ableitungen sowie geomorphologische Landformen.

2.1.5.7 Gewässer

Informationen zu Gewässerläufen wurden dem ATKIS DLM (BKG 2016) entnommen. Hierbei wurden die Distanz zum nächsten Fließgewässer im Puffer sowie die Breitenklasse (0-3, 3-6 oder 6-12 m Breite) des nächsten Fließgewässers entnommen. Zusätzlich wurden jeweils der Flächenanteil von Fließgewässern und Stehgewässern im Puffer berechnet. Zur relativen Einordnung der Position der Messstelle im Hinblick auf Fließgewässer und deren Einzugsgebietsgrenzen nach verschiedenen Strahler-Ordnungen wurde die *Multioorder hydrologic positioning Indices* (LP, SD, DSD jeweils pro Strahler-Ordnung) aus dem MOHP-Datensatz (Nölscher et al. 2022) herangezogen.

2.1.5.8 Landnutzung

2.1.5.8.1 Nicht-landwirtschaftliche Landnutzung

Allgemeine Landnutzungskategorien wurden dem Landbedeckungsmodell für Deutschland LBM-DE2018 entnommen (BKG 2018) und der prozentuale Flächenanteil im Puffer jeweils für Forst, Ackerland, Bebauung und Grünland berechnet. Flächenanteile wurden für folgende Landnutzungen entnommen:

- Deponien

- ▶ Städtische Grünflächen
- ▶ Produktionsanlagen
- ▶ Sport/Freizeitanlagen

Flächenanteile wurden für folgende Landnutzungen dem DLM (BKG 2016) entnommen:

- ▶ Kläranlagen, Kleingartenanlagen
- ▶ Freizeitanlagen
- ▶ Gewerbe/Industrie
- ▶ Summe Strecke Bahnanlagen- und Gleise im Puffer in Metern

2.1.5.8.2 Landwirtschaftliche Landnutzung

Informationen zu Menge und Zeitpunkt der tatsächlich durch die Landwirtschaft auf einzelnen Schlägen aufgebrauchten PSM stehen nicht zur Verfügung. Als Annäherung sind in erster Linie Informationen zu den angebauten Kulturen notwendig. Diese können in einem zweiten Schritt mit der Zulassungssituation verknüpft werden um die potenziell maximal flächenhaft aufgewandten PSM-Mengen abzuschätzen.

Hinsichtlich der benötigten Informationen zu den angebauten Kulturen ergab eine eingehende Recherche potenziell nutzbarer Datenquellen, dass langjährige Zeitreihen in der benötigten Auflösung nicht zur Verfügung stehen. Hochaufgelöste Fernerkundungs-Produkte, die auf Sentinel-1, Sentinel-2 und Landsat 8-Daten basieren, stehen ab 2017 (Blickensdörfer et al. 2022; Griffiths et al. 2019; Preidl et al. 2020) zur Verfügung. Von einer eigens für dieses Projekt durchgeführten Auswertung älterer Landsat 5-Daten (seit 1984 verfügbar) wurde aufgrund fehlender Ground truth-Daten („Bodenwirklichkeit“, Feldvergleich), Datenlücken aufgrund von Bewölkung sowie des unverhältnismäßigen technischen und fachlichen Aufwands für eine bundesweite Erfassung abgesehen. Agrarstatistische Erhebungen auf Gemeindeebene werden nur alle vier bis sechs Jahre (1999, 2003, 2007, 2010, 2016) bereitgestellt. Daher wurden zusätzlich Agrarantragsdaten des Integriertes Verwaltungs- und Kontrollsystem (InVeKoS) angefragt und ausgewertet.

2.1.5.8.3 Agrarantragsdaten (InVeKoS) auf Schlagebenen

Das InVeKoS (Integriertes Verwaltungs- und Kontrollsystem) „dient der Verwaltung, Überwachung und Kontrolle aller flächen- und tierbezogenen Interventionen der Gemeinsamen Agrarpolitik (GAP) (z. B. Direktzahlungen und flächen- und tierbezogene Interventionen zur Entwicklung des ländlichen Raums) durch die EU-Länder und stellt sicher, dass EU-weit umfassende und vergleichbare Daten verfügbar sind“ (EU 2024). Parzellierte, geodatenbasierte Antragsdaten erlauben eine schlaggenaue Zuordnung der beantragten Kulturen und werden jährlich erfasst. Nach Anfrage bei den zuständigen Stellen der Bundesländer konnten für die 13 Flächenbundesländer anonymisierte Antragsdaten weiterverarbeitet werden. Durch die unterschiedliche zeitliche Abdeckung der bereitgestellten Daten stehen nur drei Jahre flächendeckend zur Verfügung (2017-2019) (Anhang: Abbildung 127). Nach Bereinigung der Geometrien und dem Verschnitt mit dem 1 km-Puffer um die Messstellen (Anhang: Abbildung 128) wurden die z.T. bundesland- und jahresspezifischen Codes und Bezeichnungen der Kulturen harmonisiert und in einem einheitlichen Schlüssel zusammengefasst. InVeKoS-Codes der einzelnen Nutzungen erwiesen sich in den bereitgestellten Geodaten als nicht konsistent. Zwischen den Bundesländern als auch zwischen verschiedenen Jahren innerhalb eines Bundeslandes konnten Unterschiede zwischen dem Code und der Nutzungsbeschreibung

beobachtet werden. Daher musste ein Übersetzungsschlüssel angefertigt werden. Dieser „*Kulturschlüssel*“ entstand in Anlehnung an die Kulturklassen, die in aktuellen Fernerkundungsprodukten gewählt wurden (Blickensdörfer et al. 2022; Griffiths et al. 2019). Mais und Silomais wurden zusammengefasst. Zusätzlich wurden Klassen für sonstige landwirtschaftliche Flächen mit und ohne potenziellem Pestizideinsatz hinzugefügt. Die Flächen der Einzelschläge wurden anhand des Übersetzungsschlüssels pro Messstellen-Puffer zusammengefasst.

Das abgeleitete Datenprodukt enthält somit pro Messstelle die jährliche Fläche pro Kulturklasse im Puffer der Messstelle. Aufgrund des hohen Aufwands zur Beschaffung, Bereinigung und Harmonisierung von InVeKoS-Daten sind perspektivisch auf Fernerkundung basierende Datenprodukte zu favorisieren. Alternativ zu den aus InVeKoS extrahierten Landnutzungsinformationen wurde der auf kombinierten Zeitreihen von Sentinel-1, Sentinel-2 und Landsat 8 basierende Datensatz von Blickensdörfer et al. (Blickensdörfer et al. 2022) verwendet. Dieser steht momentan für die Jahre 2017, 2018 und 2019 zur Verfügung.

2.1.6 Weitere, nicht verwendete Daten

2.1.6.1 Geschätzte Aufwandmengen

Hersteller und Händler von PSM müssen dem Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) jährliche Absatz- und Ausfuhrmengen der in PSM verwendeten Wirkstoffe melden. In den letzten Jahren wurden bis zu etwa 35.000 t Wirkstoffe pro Jahr verkauft, wovon Herbizide etwa die Hälfte ausmachen. Aus jährlichen, bundesweiten Verkaufszahlen lassen sich tatsächlich aufgebrachte Wirkstoffmengen weder direkt ableiten noch räumlich und zeitlich näher auflösen. Die aufgewandten Mengen variieren nach Kulturen und Fruchtfolgen, der Art des Anbaus und den damit verbundenen Auflagen, die zum Teil an Standortbedingungen geknüpft sind. Zuzüglich zu den individuellen Entscheidungen der Landwirte sowie der längeren Lagerung von Produkten können Schwankungen der Witterung und des Schädlingsdrucks die tatsächlich ausgebrachten Aufwandmengen von Jahr zu Jahr beeinflussen. An zugelassene Wirkstoffe sind zulässige Höchstmengen pro Anbaufläche und Kultur mit Angaben zu den Zeitpunkten, Kulturstadien und weiteren spezifischen Indikationen geknüpft, die das BVL vorhält. Kultur- und stoffspezifische Schätzungen anhand von Umfragen an repräsentativen Betrieben wurden in den sog. PAPA-Erhebungen (Julius-Kühn-Institut, JKI) veröffentlicht, die sich grundsätzlich in der Größenordnung der bundesweiten Verkaufszahlen bewegen (2.1.6.3). Informationen über die räumliche Verteilung der Anbauflächen der Kulturen können einen ersten Hinweis auf die räumliche Verteilung der potenziell ausgebrachten Aufwandmengen liefern.

2.1.6.2 Maximale Aufwandmengen aus Zulassungssituation und InVeKoS-Flächen

Die Zulassungsdatenbank des BVL enthält monatliche Informationen zu zugelassenen Wirkstoffen, deren Anteil und Gehaltsangaben in Produkten sowie Indikationen und Kontraindikationen zur Anwendung auf spezifischen Kulturen, wie Mengen pro Fläche, Zeitpunkte und Kulturstadien.

Es wurden nur Zulassungseinträge berücksichtigt, die nicht ausschließlich für Haus- und Kleingarten (HUK) zugelassen sind. Zeitpunktangaben der Aufbringung wurden vereinheitlicht und in drei Kategorien zusammengefasst (Frühjahr-Sommer-, Herbst-Winter- und ganzjährige Anwendung). Dies geschah mithilfe der in der BVL-Datenbank verfügbaren Angaben zum Stadium der Kultur bzw. der Übersetzung aus BBCH-Codes (sofern angegeben).

Nach Verknüpfung der Stoffnamen der BVL-Datenbank und der FARM-Datenbank wurden unterschiedliche Angaben des Wirkstoffgehalts am Produkt harmonisiert. Im zweiten Schritt mussten auch Angaben der Aufwandmengen harmonisiert werden. Ziel war es, Gehaltsangaben in Gewichtsangaben und Anwendungsangaben in Gewicht pro Anbaufläche umzuwandeln, um diese Datensätze letztendlich miteinander verknüpfen zu können. Unsicherheiten ergeben sich bei Aufwandmengen, die sich auf die Behandlung von Saatgut oder die Behandlung einzelner Stämme bei Gehölzen beziehen und somit nicht sinnvoll in ein Gewicht pro Anbaufläche umgerechnet werden können. Diese wurden als im Kontext des Projekts wenig relevant angesehen und daher vernachlässigt.

Die relativ kleinteiligen Kulturangaben der BVL-Datenbank wurden in den Kulturschlüssel übersetzt, dabei wurde sich an den Zuordnungen orientiert, die bereits für die Harmonisierung der InVeKoS-Daten verwendet wurden. Der finale, vereinheitlichte Zulassungsdatensatz wurde zu jährlichen Aufwandmengen aggregiert. Durch die Vereinheitlichung der Stoffnamen sowie der Kulturzuordnungen war es dadurch möglich, die potenziell maximalen Aufwandmengen mit den Flächenangaben der jeweiligen Kulturen im Umkreis der Grundwassermessstellen zu verknüpfen.

Beim zeitlich aggregierten Datensatz wurden zuvor die Kulturflächenangaben im verfügbaren Zeitraum gemittelt sowie der Median der zugelassenen Aufwandmengen ermittelt, bevor Aufwandmengen mit Flächenangaben multipliziert wurden.

Die anhand der BVL-Zulassungsdatenbank geschätzten Aufwandmengen wurden in den in den weiteren Kapiteln vorgestellten Ergebnissen **nicht** berücksichtigt. Dies rührt daher, dass sich die Landnutzungsinformation (Fläche der Kultur im Puffer einer Messtelle) als ausreichend erwiesen hatte und aus den geschätzten Aufwandmengen kein Mehrwert für die Modelle generiert werden konnte. Dies lässt sich jedoch anhand der Auswahl an Stoffen erklären, die in den Modellen berücksichtigt wurden: Es handelt sich um Metaboliten von Wirkstoffen, die hauptsächlich einer Kultur zuzuordnen sind bzw. die Mehrheit der Anwendung sich auf eine Kultur bezieht (wenn auch zu einem geringeren Anteil auf weitere Kulturen). Daher sind die geschätzten Aufwandmengen mit den Anbauflächen stark korreliert. Nur bei Stoffen, deren Zulassung sich nennenswert auf mehrere Kulturen und somit auch in signifikant unterschiedliche Aufwandmengen pro Hektar verteilt, wäre ein Mehrwert dieser Schätzung gegenüber der Kulturfläche zu erwarten gewesen. Für besagte Stoffe konnten allerdings aufgrund weiterer Limitationen (siehe Kapitel 2.2.1 und 2.2.3.6) keine robusten Modelle entwickelt werden. Weiterhin ist zu bedenken, dass gerade bei Altstoffen (d.h. Wirkstoffen mit langer Anwendungshistorie und ihre Metaboliten) in der Vergangenheit z.T. höhere Aufwandmengen zugelassen waren als heutzutage, die Zulassungsdatenbank diese jedoch nicht mehr beinhaltet.

2.1.6.3 Geschätzte PSM-Aufwandmengen aus PAPA-Erhebungen

PAPA steht für das *Panel Pflanzenschutzmittel-Anwendungen*. Seit 2011 werden für eine Erhebung des Julius-Kühn-Instituts (JKI) kulturspezifische Netze von Erhebungsbetrieben gebeten, jährliche PSM-Anwendungsdaten detailliert zu erfassen und in anonymisierter Form dem JKI zur Verfügung zu stellen. Dadurch konnten für relevante Kulturen anhand dieser repräsentativen Betriebe bundesweite Aussagen zur Pflanzenschutzanwendung abgeleitet werden (PAPA 2024). Diese Schätzungen werden in Cooke et al. (2024) verwendet, allerdings nicht in den in weiteren Kapiteln beschriebenen Modellen.

Jeweils pro Stoff, Kultur und Jahr stehen somit zur Verfügung:

- Mittlere Behandlungsfläche in ha

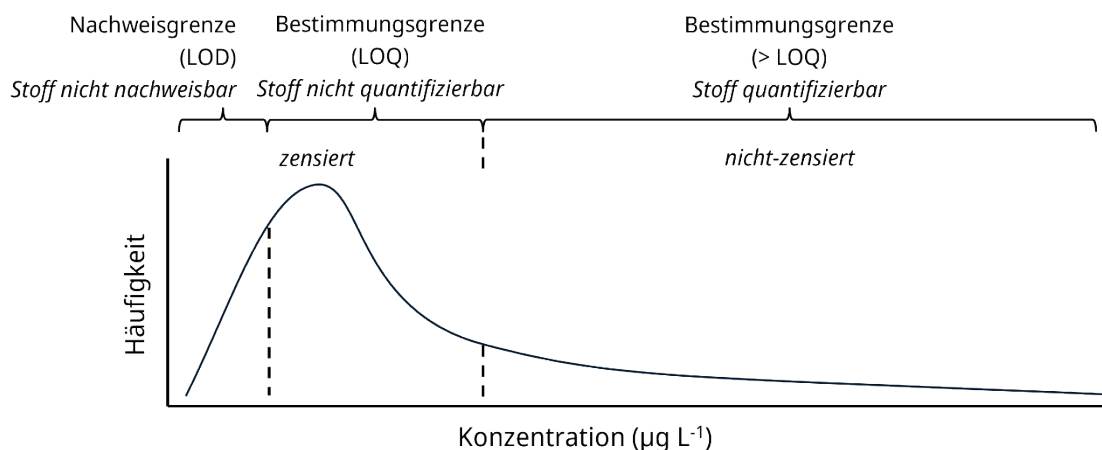
- ▶ Mittlere Aufwandmenge kg ha^{-1}
- ▶ Unteres Konfidenzintervall der Behandlungsfläche in ha
- ▶ Oberes Konfidenzintervall der Behandlungsfläche in ha
- ▶ Relativer Anteil der Anzahl von Erhebungsbetrieben, für die ein Stoff pro Kultur und Jahr angewandt wurde
- ▶ Relativer Anteil am Wirkstoffbereich (bezogen auf Behandlungsindex-Werte)

2.2 Verwendete Methoden für die statistische Auswertung

2.2.1 Methoden für linkszensierte Daten

Zur Bestimmung der Konzentrationen von Stoffen werden Verfahren wie Flüssigchromatographie mit Massenspektrometrie-Kopplung (LC-MS/MS) verwendet. Für dieses Messverfahren wird jeweils eine substanzspezifische analytische Nachweisgrenze (LOD; engl.: limit of detection) und eine Bestimmungsgrenze (LOQ; engl.: limit of quantification) statistisch definiert. Werte oberhalb der LOQ sind quantifizierte Messwerte, während Messungen unterhalb der LOQ nicht mit ausreichender Präzision bestimmt werden können. Messungen oberhalb der LOD, aber unterhalb der LOQ, sind Nachweise, die sich nicht verlässlich genug quantifizieren lassen. Sie werden daher als „kleiner als“ angegeben und werden als linkszensiert bezeichnet (Abbildung 1). Die LOQ sollte mindestens 30% des regulatorischen Grenzwertes ($0,03 \mu\text{g L}^{-1}$ im Falle von $0,1 \mu\text{g L}^{-1}$) betragen.

Abbildung 1: Beispiel einer Verteilung von Konzentrationen mit linkszensiertem Anteil unterhalb der Nachweis- (LOD) oder Bestimmungsgrenze (LOQ).



Quelle: eigene Darstellung BGR

Im Gesamtdatensatz liegen 96,6% aller Messungen (unabhängig von Stoff oder Messstelle) unterhalb der jeweiligen LOQ. Die Werte der LOQ liegen zwischen $0,001 \text{ ng L}^{-1}$ und $1 \mu\text{g L}^{-1}$. Pro Stoff wurden teilweise mehrere LOQ angewandt, mit bis zu 48 unterschiedlichen Werten zwischen $0,02 \text{ ng L}^{-1}$ und $0,1 \mu\text{g L}^{-1}$ (Desethylatrazin). Die Heterogenität der LOQ zwischen Messstellen, Stoffen, Bundesländern und über die Zeit erschwert räumliche Vergleiche der Fundsituation und die Analyse von zeitlichen Trends.

Etwa 12% der zensierten Messungen sind als unterhalb der LOD ausgewiesen. Dateneinträge, die als $< LOD$ angegeben wurden, stammen ausschließlich aus drei Bundesländern. Dies umfasst 4.135 Messstellen in diesen Bundesländern, für die mindestens eine Messung als $< LOD$ angegeben wurde. Es ist davon auszugehen, dass Informationen über die jeweilige Nachweisgrenze in den Bundesländern unterschiedlich gehandhabt werden. Ob Messwerte unterhalb der LOQ auch als oberhalb der LOD zu verstehen sind, ist nicht zweifelsfrei aufzulösen. In den nachfolgenden Analysen wurde davon ausgegangen, dass „ $< LOQ$ “ nicht auch „ $< LOD$ “ bedeutet. Für Fälle, in denen sowohl die LOD als auch die LOQ angegeben wurde, die Messung aber als „ $< LOQ$ “ gekennzeichnet wurde, wurde ausschließlich die LOQ berücksichtigt. Die LOD könnte als untere Grenze zur Abschätzung des Intervalls, in das eine zensierte Messung fallen würde, dienen. Angaben zur LOD wurden in einem Großteil der Daten nicht mitgeliefert, wären jedoch hilfreich.

Messungen mit einer sehr hohen LOQ ($1 \mu\text{g L}^{-1}$) stellen einen sehr geringen Anteil am Datensatz dar und wurden aufgrund von Zweifeln an der Datenqualität nicht berücksichtigt. Die höchste gemeinsame LOQ wurde auf $0,1 \mu\text{g L}^{-1}$ festgelegt. Somit lassen sich die Häufigkeiten der Überschreitungen der $0,1 \mu\text{g L}^{-1}$ im Datensatz leicht vergleichen.

Für einen Teil der Analysen wurde eine Konzentrationseinteilung vorgenommen, die in a) $< LOD$, b) $< LOQ$, c) $< 0,1 \mu\text{g L}^{-1}$, d) $> 0,1$ und $< 1 \mu\text{g L}^{-1}$ und in e) $> 1 \mu\text{g L}^{-1}$ einteilt. Ein Teil der Messungen in Fall b) könnte auch zu Fall a) gehören. Aufgrund der heterogenen, bis an $0,1 \mu\text{g L}^{-1}$ heranreichenden LOD-Werte, kann eine scharfe Trennung zwischen Fall b) und c) nicht gewährleistet werden. Allein die Fälle d) und e) lassen sich klar von den anderen Fällen abtrennen. Trotz dieses Fehlers ist diese Art der Darstellung für die Fragestellungen in diesem Projekt ausreichend. So werden Messungen unterhalb der LOD in vielen Auswertungen als eine Gruppe aufgeführt (LAWA 2024).

Die höchste gemeinsame Bestimmungsgrenze von $0,1 \mu\text{g L}^{-1}$ ist gleichzeitig der Grenzwert für Wirkstoffe und relevante Metaboliten und somit relativ hoch, sodass eine differenziertere Betrachtung der Messungen unterhalb dieses Wertes wünschenswert wäre. Daraus ergibt sich eine Notwendigkeit, nach Möglichkeit auch Aussagen über den Teil der Daten unterhalb der Bestimmungsgrenze treffen zu können.

Zensierte Daten werden in der Wissenschaft unterschiedlich gehandhabt. Die einfachsten Methoden substituieren die zensierten Werte mit Null, der Hälfte oder einem Viertel der LOQ. Gerade das Nullsetzen ist als problematisch zu erachten, wenn bei Werten oberhalb der LOD zwar von nicht quantifizierbaren, aber dennoch nachweisbaren Konzentrationen gesprochen wird. Substitution kann je nach Datenlage Artefakte und Verzerrungen hinzufügen bzw. vorhandene Zusammenhänge und Trends maskieren. Gerade in Fällen mit einer großen Bandbreite an unterschiedlichen LOQ kann Substitution problematisch sein. Zum korrekteren Umgang mit zensierten Daten werden je nach Anteil der Zensur und Datenmenge verschiedene Verfahren empfohlen. Dazu zählen parametrische und nicht-parametrische Methoden, die die Eigenschaften der Gesamtverteilung anhand der nicht-zensierten Daten schätzen. Dies kann mithilfe der Maximum-Likelihood-Estimation (MLE), dem Kaplan-Maier-Schätzer (KM) oder den Regression-on-Order-Statistics (ROS) geschehen. Für zahlreiche statistische Auswertungen existieren analoge Versionen, die mit zensierten Daten umgehen können. Die Verwendbarkeit und Robustheit einzelner Verfahren ist dabei von der Anzahl der insgesamt vorhandenen Datenpunkte als auch vom Anteil zensierter Daten abhängig. Die meisten Methoden werden für einen Zensurgrad von unter 80% empfohlen, da darüber hinaus zu viel Unsicherheit bei der Schätzung der Verteilung besteht (Helsel 2012).

In vielen Fällen zeigen Stoffkonzentrationen in der Umwelt eine rechtsschiefe Verteilung (Helsel 2012). Um die Anzahl an Überschreitungen von Schwellenwerten kleiner als $0,1 \mu\text{g L}^{-1}$ zu berechnen, muss eine Schätzung der Verteilung der Verteilungsparameter durchgeführt werden. Wird eine oft angenommene logarithmische Normalverteilung vorausgesetzt, dann lässt sich basierend auf der kumulativen Dichtefunktion ein Anteil p_{exc} der Daten, der oberhalb eines gewählten Schwellenwertes L liegt, gemäß Formel 1 schätzen. D.h. es wird die Überschreitungswahrscheinlichkeit, dass eine lognormalverteilte Zufallsvariable X den Schwellenwertes L übersteigt, gesucht. Dabei ist Φ die kumulative Verteilungsfunktion der Standardnormalverteilung, die Standardabweichung σ und der Mittelwert μ sind die Momente der zugrundeliegenden Normalverteilung von $\ln X$, wobei X lognormalverteilt ist.

Formel 1: Berechnung des Verteilungsparameters p_{exc} .

$$p_{exc} = 1 - \Phi \left[\frac{\ln L - \mu}{\sigma} \right]$$

Angesichts der Datenlage wurde ein Schwellenwert angestrebt der niedriger ist als $0,1 \mu\text{g L}^{-1}$, um auch eine Betrachtung von geringeren Konzentrationen zu ermöglichen. In Cooke et al. (Cooke et al. 2024) werden Überschreitungen anhand des Schwellenwertes $0,05 \mu\text{g L}^{-1}$ als sinnvoll erachtet, da dieser Wert zu niedrig und dadurch robust, da auch Messwerte oberhalb der LOQ miteinbezogen wurden und nicht ausschließlich zensierte Daten.

Zusammenfassend werden in den Auswertungen die folgenden Möglichkeiten verwendet, die Wirkstoff- bzw. Metabolitenkonzentration als Zielvariable in verschiedenen Analysen zu erfassen und darzustellen:

- ▶ Einteilung in fünf Bereiche: a) $< \text{LOD}$, b) $< \text{LOQ}$, c) $< 0,1$, d) $> 0,1$ und $< 1 \mu\text{g L}^{-1}$ und in e) $> 1 \mu\text{g L}^{-1}$ einteilt. Dies hat den Vorteil, dass die Daten nicht verändert wurden. Die Bandbreite der LOQ und LOD sorgt allerdings dafür, dass Klassen a-c nicht trennscharf unterschieden werden können.
- ▶ Berechnung von Überschreitungshäufigkeiten: Für Schwellenwerte von $0,1 \mu\text{g L}^{-1}$ und höher ist dies unproblematisch, für einen Schwellenwert von $0,05 \mu\text{g L}^{-1}$ handelt es sich um geschätzte Überschreitungen. Diese Schätzung ist nur für eine begrenzte Anzahl Stoffe mit einem relativ geringen Zensuranteil sinnvoll.

2.2.2 Explorative Datenanalyse

Um einen Überblick über die übermittelten Messdaten zu erhalten, wurden die verfügbare Anzahl an Messstellen und Länge der Messreihen (auch hinsichtlich der Unterschiede zwischen den Bundesländern), sowie das Stoffspektrum untersucht (3.1). In einem weiteren Schritt wurden die Fundsituation stoffunabhängig und stoffspezifisch mit Stoffeigenschaften und Standortfaktoren in Verbindung gebracht. Weiterhin wurde die Fundsituation auf räumliche und zeitliche Muster und Trends geprüft (3.1.3).

Für Korrelationen zwischen Standortfaktoren und Stoffkonzentrationen wurde der Rangkorrelationskoeffizient nach Spearman ermittelt (Spearman 1904). Kontinuierliche Variablen wurden unverändert einbezogen, kategorische Variablen wurden „one-hot“ kodiert, d.h. dass einzelne Ausprägungen (Kategorien) dieser Variable als eigene Variable gehandhabt wurden und ihnen jeweils die Werte 0 bzw. 1 (nicht vorhanden/ vorhanden) zugewiesen wurden. Zusammenhänge zwischen ausschließlich kategorischen Variablen wurden zusätzlich mittels Cramer's V untersucht, ohne eine Kodierung vornehmen zu müssen (Cramér 1946).

2.2.3 Methoden des maschinellen Lernens

2.2.3.1 Was ist maschinelles Lernen?

Das maschinelle Lernen ist Teil des Gebiets der Künstlichen Intelligenz und umfasst Methoden, bei denen Computeralgorithmen aus Daten Modelle entwickeln, die u.a. zur Mustererkennung, Entscheidungsfindung und Vorhersage genutzt werden können. Die zu erlernenden Regeln und Zusammenhänge werden dabei nicht wie bei der konventionellen Modellierung zuvor festgelegt und fest einprogrammiert, sondern werden explizit aus den Daten entnommen und daher „gelernt“. Methoden des maschinellen Lernens können insbesondere bei großen und unstrukturierten Datenmengen ihre Stärken im Erkennen von vorher unbekanntem Zusammenhängen aufzeigen. Der Lernprozess wird als Training bezeichnet und beinhaltet ein iteratives Anpassen von Modellparametern sowie den Abgleich von Modellvorhersagen anhand der Ausgangsdaten. Dabei wird eine anwendungsspezifische Verlustfunktion (engl.: loss function) optimiert, sodass der Fehler, d.h. die Abweichung zwischen Modellvorhersage und Daten, reduziert wird. Der Lernprozess kann mittels spezifischer Einstellungen – sogenannter Hyperparameter – gesteuert werden, die üblicherweise während des Trainingsablaufs optimiert werden müssen. Eine Evaluierung der Modellgüte geschieht anhand von ausgewählten Metriken und wird anhand eines zuvor abgespalteten Datenpakets (Testdaten) durchgeführt, das üblicherweise 10-30% der Gesamtdatenmenge umfasst. Da der Algorithmus diese Testdaten für die Modellentwicklung nicht verwendet hat, wird somit eine unabhängige Modellevaluierung ermöglicht. In der Praxis wird häufig nicht nur eine einmalige Aufteilung in Trainings- und Testdaten, sondern eine Kreuzvalidierung anhand mehrerer Splits durchgeführt (siehe Abschnitt 2.2.3.4).

Abbildung 2: Vereinfachtes Schema zur Einteilung der Daten beim Maschinellen Lernen.



Quelle: eigene Darstellung BGR

Train-Set: Modell lernt Zusammenhänge anhand von Trainingsdaten; Validation-Set: Wird genutzt um Hyperparameter zu optimieren; Test-Set: während des Trainings ungesehener Datensatz wird zur Evaluierung der Performance genutzt.

Das maschinelle Lernen lässt sich in überwachtes und unüberwachtes Lernen einteilen (engl.: supervised, unsupervised learning). Beim unüberwachten Lernen gibt es keinen zuvor definierten Output. Das Modell ordnet die Daten anhand ihrer inneren Struktur selbst und sortiert sie in selbstgewählte Gruppen (engl.: Cluster), was u.a. zur Dimensionsreduktion genutzt werden kann. Beim überwachten Lernen liegen Werte für die Zielvariable bereits vor, welcher eine kontinuierliche (Regression) oder eine kategorische Zielvariable (Klassifikation) sein kann. Weitere Methoden sind u.a. das halbüberwachte Lernen, selbstüberwachtes Lernen oder das Reinforcement Learning. Im Kontext dieser Arbeit wurde Klassifikation (überwachtes Lernen) anhand von Konzentrationsklassen (2- bzw. 3-Klasseneinteilung) durchgeführt.

2.2.3.2 Räumliches vs. Zeitliches Modell

Aufgrund der heterogenen Länge der Messzeitreihen für zahlreiche Stoffe, wurden zwei Modellansätze verfolgt: Ein vorwiegend räumliches Modell, das statische und zeitlich aggregierte Features und Zielvariablen verwendet, sowie ein räumlich-zeitliches Modell. Ein „hybrides“ Modell, welches das Jahr der Messung als Feature enthält, wurde ebenfalls getestet. Zeitliche Features wurden für das räumliche Modell für den Zeitraum ab 2010 aggregiert, ebenso geschah die Zuordnung der Konzentrationsklassen anhand des zeitlichen Medians der Konzentrationen. Räumliche Modelle wurden mit Random-Forest und Neuronalen Netzen

trainiert, das räumlich-zeitliche Modell wurde nur mit einem Neuronalem Netz durchgeführt (siehe 2.2.3.4). Eine Einschätzung zur zeitlichen Datenverfügbarkeit wird in 3.1.3.3 aufgeführt.

2.2.3.3 Modellgüte

Zur Evaluierung der Modellgüte stehen zahlreiche Metriken zur Verfügung, die entsprechend der Aufgabe (Regression bzw. Klassifikation) gewählt werden können und je nach Fragestellung Vor- und Nachteile mit sich bringen. In Tabelle 2 wird eine klassische Konfusionsmatrix (engl.: confusion matrix) für ein Klassifikationsmodell dargestellt. Das Modell sagt für jeden Datenpunkt eine Klasse voraus. Nach Abgleich der Modellvorhersage mit den tatsächlichen Daten werden vier Fälle unterschieden: richtig positive (true positive, TP), falsch positive (false positive, FP), richtig negative (true negative, TN) und falsch negative (false negative, FN) Vorhersagen. Die Konfusionsmatrix liefert die absolute Anzahl an Datenpunkten, die jeweils den Fällen zugeordnet werden kann und dient somit als Maß dafür wie gut verschiedene Klassen vorhergesagt werden. Es ist zu beachten, dass jede der Klassen unabhängig voneinander als positiv definiert werden kann, um die Performance für alle Klassen separat zu analysieren. Bei einem binären Klassifikationsmodell (2 Klassen) ist jeweils die Klasse, die nicht als positiv definiert ist, automatisch die negative Klasse. Bei Mehrklassen-Klassifikationen (z.B. 3 Klassen) wird der "One-vs-Rest"-Ansatz verwendet, indem die jeweilige Klasse als positiv und alle anderen Klassen als negativ behandelt werden, sodass für jede Klasse eine separate 2x2-Konfusionsmatrix entsteht. Die Umwandlung der Konfusionsmatrizen in One-vs-Rest-Konfusionsmatrizen ist beispielhaft für eine binäre Klassifikation und für eine Mehrklassen-Klassifikation (3 Klassen) in Abbildung 129 und Abbildung 130 in Anhang A.5 dargestellt.

Tabelle 2: Bei Klassifikationsproblemen können bei Abgleich der vorhergesagten mit der tatsächlichen Klasse die vier möglichen Fälle TP, FN, FP und TN dargestellt werden.

Kategorie	Bezeichnung
Tatsächliche Klasse (+) und Vorhergesagte Klasse (+)	Richtig positiv (TP)
Tatsächliche Klasse (-) und Vorhergesagte Klasse (-)	Richtig negativ (TN)
Tatsächliche Klasse (-) und Vorhergesagte Klasse (+)	Falsch positiv (FP)
Tatsächliche Klasse (+) und Vorhergesagte Klasse (-)	Falsch negativ (FN)

Als simpelste und intuitivste Metrik kann die Accuracy dienen (Exaktheit, Formel 2), welche sich direkt aus dem Anteil korrekt vorhergesagten Daten ergibt.

Formel 2: Berechnung der Accuracy zur Bewertung des ML-Modells.

$$Accuracy_i = \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}$$

Wobei i die als positiv definierte Klasse darstellt. Die Accuracy kann die Modellgüte bei stark unbalanzierten Daten (deutlich mehr Fälle innerhalb einer Klasse im Vergleich zu anderen Klassen) jedoch deutlich überschätzen. Wenn sich beispielsweise 90% der Fälle einer Klasse zuordnen lassen, kann das Modell eine Accuracy von 0,9 erreichen, selbst wenn alle Fälle einer zweiten Klasse inkorrekt vorhergesagt wurden. Eine weitere Metrik ist die *Precision* (Genauigkeit), welche steigt, sobald der Anteil der FP minimiert wird (Formel 3):

Formel 3: Berechnung der Precision zur Bewertung des ML-Modells.

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$

Praktisch heißt das, dass die Optimierung der Precision die Reduktion des Anteils an Messstellen anstrebt, an denen die Konzentrationen fälschlicherweise überschätzt wurden. Der *Recall* (auch *Sensitivity*; Vollständigkeit) steigt dagegen, sobald der Anteil an FN minimiert wird (Formel 4):

Formel 4: Berechnung des Recall zur Bewertung des ML-Modells.

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

Dementsprechend liegt der Fokus darauf, möglichst wenige Kontaminationen zu unterschätzen. Je nach Zielsetzung können sowohl eine hohe Precision als auch ein hoher Recall von Interesse sein. Als Abwägung beider Metriken kann der F1-Score, der sich aus dem harmonischen Mittel beider Werte berechnet, als kombinierte Metrik genutzt werden (Formel 5):

Formel 5: Berechnung der kombinierten Metrik F1-Score zur Bewertung des ML-Modells.

$$F1_i = 2 * \frac{Precision_i * Recall_i}{Precision_i + Recall_i}$$

Die Werte des F1-Scores sind dabei gleichermaßen von den Werten des Recalls und der Precision abhängig. Die genannten Metriken können separat für jede Klasse *i* als auch für das Gesamtmodell berechnet werden und dabei mehr als nur zwei Klassen betrachten. Beim F1-Score kann der Macro-F1-Score klassenübergreifend verwendet werden, wobei jede Klasse gleich gewichtet wird (Formel 6):

Formel 6: Berechnung der kombinierten Metrik Macro-F1-Score zur Bewertung des ML-Modells.

$$F1_{macro} = \frac{1}{K} \sum_{i=1}^K F1_i$$

Die klassenspezifischen Berechnungen von Precision, Recall und F1-Score sowie die Ermittlung des Macro-F1-Scores sind beispielhaft für eine binäre Klassifikation (2 Klassen) und für eine 3-Klassen-Einteilung in Abbildung 129 und Abbildung 130 in Anhang A.5 dargestellt. Der Macro-F1-Score wird oft bei stark unbalancierten Daten verwendet, hat jedoch den Nachteil, dass es zu einer stärkeren Repräsentation der Minderheitsklasse kommen kann (problematisch bei sehr kleinen Klassen). Es gibt eine Vielzahl weiterer Metriken, die in diesem Bericht jedoch nicht behandelt werden. Hierzu wäre die *Area under curve* der *Receiver-Operating-Curve* (ROC AUC) zu nennen, welche die Abwägung zwischen der *True-positive-Rate* (TPR) und der *False-Positive-Rate* (FPR) darstellt. Auch die Fläche unterhalb der Precision-Recall-Kurve (PR AUC) wird verwendet.

2.2.3.4 Im Projekt verwendete Algorithmen

In diesem Projekt wurden mithilfe von Random-Forest-Modellen (RF) und Neuronale Netzen (NN) für verschiedene Stoffmodelle verwendet. Sowohl die Modelloptimierung, als auch die Evaluierung der Modellgüte erfolgte anhand des Macro-F1-Scores. Zum weiteren Vergleich wurden Recall, Precision und F1-Score jeweils für jede Klasse einzeln berechnet. Die Robustheit von Modellen sowie von ermittelten *Feature Importances* (Merkmalsbedeutung) wurde

bewertet, indem die Abhängigkeit von zufälliger Datenpartitionierung (RF, NN) durch Verwendung von unterschiedlichen *Seeds*¹ (Startbedingungen) und verschiedenen Initialisierungen (NN) getestet wurden. Zum Vergleich der Modellgüte wurde ein sog. *Benchmark*-Modell trainiert, d.h. ein einfacheres Basismodell, das als Maßstab hinzugezogen wurde.

Random-Forest

Random-Forest ist eine nichtparametrische Ensemblemethode (Breiman 2001), bei der eine hohe Anzahl (z.B. 1000) möglichst unkorrelierter Entscheidungsbäume erzeugt wird. Die Robustheit der Methode basiert auf der Aggregation individueller Vorhersagen von unterschiedlichen Entscheidungsbäumen, für die pro Entscheidungssplit (Knoten) nur eine begrenzte, zufällig ausgewählte Teilmenge der zur Verfügung stehenden Features verwendet wird. Darüber hinaus werden die einzelnen Bäume basierend auf zufälligen Stichproben aus dem Datensatz erstellt. Die Aggregation der Modellvorhersagen der einzelnen Bäume erfolgt durch Berechnung des arithmetischen Mittels (für Regressionsprobleme), bzw. durch Ermittlung der mehrheitlich gewählten Klasse (für Klassifikationsprobleme).

Die Entwicklung der Random-Forest-Modelle erfolgte mithilfe der R-Pakete *ranger* (Wright et al. 2017) und *mlr3* (Lang et al. 2019). Dabei wurden probabilistische Modelle (Malley et al. 2012) unter Anwendung des *Gini-Index* als Splitting-Regel erstellt. Der Algorithmus umfasst mehrere Hyperparameter, die gemäß verschiedener Tuning-Strategien optimiert werden können. Für eine Vielzahl an Problemstellungen ist die Anzahl der zufällig ausgewählten Variablen, die pro Entscheidungssplit berücksichtigt werden, *mtry*, am sensitivsten (Probst et al. 2019). In diesem Projekt wurde der Hyperparameter *mtry* daher mit Hilfe einer Gittersuche (engl.: *grid search*) optimiert, während ein weiterer Hyperparameter, die Anzahl der Bäume (*num.trees*) nach einer Sensitivitätsanalyse auf einen Wert von 1000 gesetzt wurde. Oftmals wird ein voreingestellter Wert von 100 verwendet. Modelle mit mehr Bäumen sind meist deutlich robuster, sofern die dafür benötigte Rechenleistung vorhanden ist.

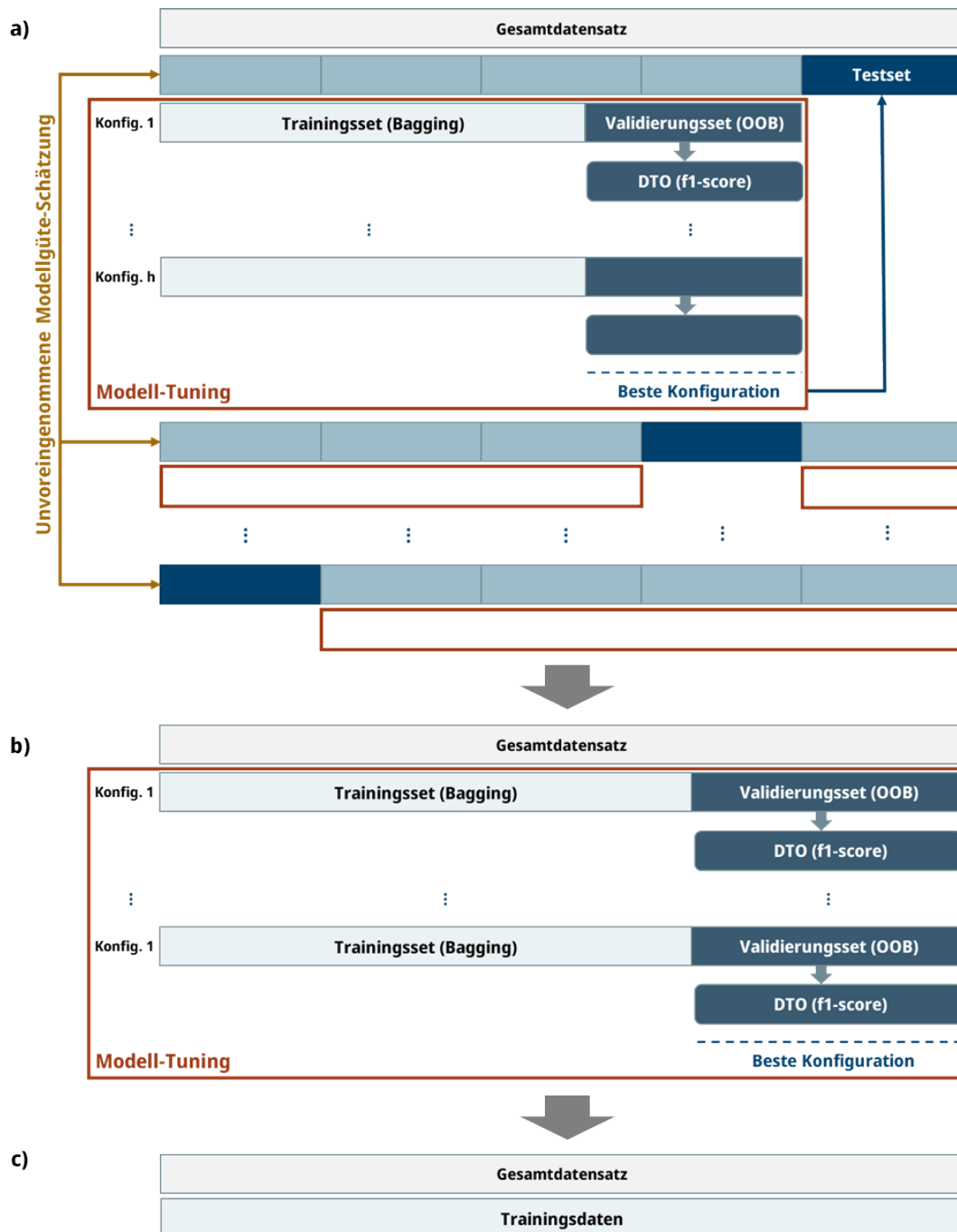
Bei der finalen Modellauswahl des Hyperparameter-Tunings wird das beste Modell gemäß der gewählten Modellgütemaße ausgewählt. Unabhängig von den gewählten Algorithmen, ist bei der Modellauswahl die Vermeidung von *Overfitting* (Überanpassung), als auch *Underfitting* (Unteranpassung) relevant. *Overfitting* tritt auf, sobald das Modell sich auf sehr spezifische Eigenschaften der Daten konzentriert, die ggf. als Rauschen bezeichnet werden können. Solche overfittenden Modelle erzielen oft eine hohe Performance anhand der Trainingsdaten, können ungesehene Testdaten jedoch nur schlecht vorhersagen, also schlecht generalisieren. Dies wirkt sich darin aus, dass die Modelle z.T. irrelevante Einzelheiten der Trainingsdaten erlernen, allerdings grundsätzliche Zusammenhänge in den Daten, die zur Vorhersage der Testdaten erforderlich wären, dadurch weniger Einfluss bekommen. *Underfitting* wiederum zeigt sich durch eine bereits geringe Trainingsperformance und eine niedrige Testperformance, die unterhalb der möglichen Performance zurückbleibt. Dem Modell fehlt der nötige Grad an Komplexität, um die Daten sinnvoll abzubilden. Der Hyperparameterraum wurde eingegrenzt, um zu starkes *Overfitting* zu vermeiden. Aus fachlichen Gründen wurde eine *min.node.size* von 1 (einzelne Messstellen) als zu kleinteilig eingestuft. Um Überanpassung (*Overfitting*) zu reduzieren und die Interpretierbarkeit der Modelle zu erhöhen, wurde die minimale Anzahl an

¹ Werte, die zur Initialisierung eines Zufallszahlengenerators verwendet werden, werden als „Seeds“ bezeichnet. Durch das Festlegen eines Seeds wird sichergestellt, dass stochastische Prozesse während des Trainings und der Datenaufbereitung (wie Aufteilung oder Durchmischung der Daten) bei jeder Ausführung des Codes die gleichen Ergebnisse liefern. Die Nennung der verwendeten Seeds stellt eine Voraussetzung zur Reproduzierbarkeit der Trainingsergebnisse dar.

Beobachtungen in einem terminalen Knoten, min.node.size, auf einen festen Wert von 10 gesetzt.

Die robuste Evaluierung der Modelle erfolgte basierend auf einer 5-fachen Kreuzvalidierung (engl.: 5-fold cross-validation), kombiniert mit Bagging (bootstrap aggregating) zur Hyperparameteroptimierung (HPO). Dafür wurde der Datensatz zunächst in fünf gleichgroße Teildatensätze aufgeteilt (äußere Teildatensätze; engl.: outer folds), wobei in jeder Iteration der Kreuzvalidierung jeweils ein Teildatensatz als Testset und die restlichen Teildatensätze als Trainingsset verwendet wurden. Für jedes (äußere) Trainingsset wurde eine separate HPO basierend auf dem für Random-Forest weit verbreiteten Bootstrapping-Verfahren durchgeführt. Dabei wird für jeden Baum im Random-Forest eine unabhängige Stichprobe erzeugt, indem ein zufälliger Teil der Daten durch wiederholtes Ziehen mit Zurücklegen (Bootstrapping) ausgewählt wird. Diese Stichprobe wird für das Trainieren des jeweiligen Modells verwendet. Die nicht in der Stichprobe enthaltenen Datenpunkte werden als Out-Of-Bag (OOB) bezeichnet und können als zuverlässiges Validierungsset zur HPO verwendet werden. Standardmäßig werden für probabilistische binäre Klassifikationsmodelle ein Entscheidungsschwellenwert von 0,5 angewendet, sodass ein Datenpunkt der Klasse zugeordnet wird, die von mehr als 50% der Bäume eines Random-Forest vorhergesagt wurde. Bei stark unausgeglichene Daten führt dies jedoch häufig zur Bevorzugung der Mehrheitsklasse, sodass im Anschluss an das Modelltraining für jede Hyperparameter-Kombination eine Optimierung des Entscheidungsschwellenwertes durchgeführt wurde. Dabei wurde der Schwellenwert, ab dem ein Datenpunkt in eine bestimmte Klasse eingeordnet wird, so angepasst, dass die Güte des Modells in Bezug auf den F1-Score optimiert wurde. Für jedes äußere Trainingsset wurde das Modell mit der optimalen Kombination aus Hyperparametern und Entscheidungsschwellenwert auf dem jeweiligen Testset zur unvoreingenommenen Schätzung der Modellgüte bewertet. Eine schematische Darstellung des Modellaufbaus findet sich in Abbildung 3a. Um die optimale Hyperparameter-Kombination für den gesamten Datensatz zu erhalten, wurde Bagging auf alle verfügbaren Daten angewendet. Dabei wurden die HPO sowie die anschließende Optimierung des Entscheidungsschwellenwertes wie zuvor beschrieben für die neue Datenpartitionierung durchgeführt (Abbildung 3b). Abschließend wurde das Modell mit der besten Hyperparameter-Kombination und dem besten Entscheidungsschwellenwert auf den gesamten Daten trainiert (Abbildung 3c).

Abbildung 3: Schematische Darstellung der Modellarchitektur, die für die Random Forest (RF)-Modelle verwendet wurde. Das Modell-Tuning erfolgte anhand von Bagging unter der Verwendung des OOB als Validierungsset (innere Umrandungen). Die Evaluierung der Modelle mit optimaler Hyperparameter-Kombination und Entscheidungsschwellenwert erfolgte anhand einer 5-fachen Kreuzvalidierung jeweils auf einem unabhängigen Testset (a). Anschließend wurde das Modell-Tuning für den gesamten Datensatz wiederholt (b) und das Modell mit der besten Hyperparameter-Kombination und dem besten Entscheidungsschwellenwert auf den gesamten Daten trainiert (c).



Quelle: eigene Darstellung BGR.

Neuronales Netz

Neuronale Netze sind eine Methode des maschinellen Lernens aus dem Bereich des *Deep Learning*. Als neuronale Netze werden Modellstrukturen bezeichnet, bei denen zahlreiche Knotenpunkte (Neuronen) miteinander verknüpft sind und Signale (Daten) innerhalb der Knotenpunkte verarbeitet und weitergeleitet werden.

Ein neuronales Netz besteht typischerweise aus mehreren Schichten: Einer Input-Schicht, in die die Rohdaten eingespeist werden, mehreren *hidden layers*, in denen die Daten verarbeitet werden, und einer Output-Schicht, die eine Voraussage ausgibt. Jede Schicht verfügt über eine bestimmte Anzahl an Neuronen. Je mehr hidden layers und je mehr Neuronen pro Schicht das Netz beinhaltet, desto komplexere Zusammenhänge können gelernt werden. Allerdings steigt damit auch das Potential für Overfitting sowie die Rechenzeit.

Für das Trainieren Neuronaler Netze wird neben Trainings- und Testset ein Validierungsset benötigt (typischerweise vom Trainingsset abgespaltene Daten). Im Trainingsprozess wird das Trainingsset stückweise in das Modell gegeben (bestimmt durch den Hyperparameter batch size, der die Anzahl der Daten festlegt, die gleichzeitig vom neuronalen Netz verarbeitet werden). Jedes Neuron empfängt Daten und verarbeitet sie mit der sogenannten Aktivierungsfunktion, die durch (zufällig initialisierte) weights gesteuert wird. Die Ausgaben aus der Aktivierungsfunktion dienen dann wiederum Neuronen der nächsten Schicht als Input. Schließlich werden in der Ausgabeschicht die Outputs des letzten hidden layer zu einem Wahrscheinlichkeitswert umgewandelt, der die Zugehörigkeit zu einer bestimmten Klasse angibt. Die Güte dieser Vorhersage wird mit der Verlustfunktion anhand des Validierungssets geprüft. Daraufhin werden die weights der einzelnen Neuronen aktualisiert, um im anschließenden Lauf (Epoche) eine bessere Performance zu erzielen bzw. die Verlustfunktion im Laufe mehrerer Epochen zu minimieren. Die Aktualisierung der weights wird durch einen sogenannten optimizer durchgeführt. Der optimizer beinhaltet eine learning rate, die festlegt, wie stark die weights verändert werden. Sobald die loss function nicht mehr weiter reduziert wird, oder eine gewählte maximale Anzahl an Epochen erreicht ist, ist der Trainingsprozess abgeschlossen. Die tatsächliche Performance des Modells kann dann (wie beim Random-Forest) anhand des Testsets bewertet werden.

Die Wahl der Hyperparameter ist entscheidend für die Performance des Modells. Allerdings gibt es keine universale Konfiguration, die für alle Anwendungen optimal ist. In diesem Optimierungsprozess werden die Hyperparameter variiert und die Performance der jeweiligen Modelle verglichen (Hyperparameter-Optimierung; siehe folgender Absatz). Die Hyperparameter für ein neuronales Netz, die betrachtet wurden, sowie ihr Einfluss auf das Modell sind im Folgenden aufgelistet:

- ▶ **Batch size:** bestimmt, wie viele Trainingsdaten gleichzeitig ins Modell eingespeist werden. Eine geringe batch size kann für eine Minimierung der Verlustfunktion innerhalb weniger Epochen sorgen (Das et al. 2016; Keskar et al. 2017). Eine hohe batch size macht den Rechenprozess effektiver, da mehr Daten auf einmal verarbeitet werden (Goyal et al. 2018; You et al. 2017).
- ▶ **Maximale Anzahl an Epochen:** gibt vor, über wie viele Epochen die weights maximal aktualisiert werden, um die Verlustfunktion zu optimieren. Zeigt die Verlustfunktion bei weniger Epochen keine Verbesserung, wird der Trainingsprozess früher beendet. Die maximale Anzahl an Epochen kann übermäßig lange Trainingszeiten und Overfitting reduzieren; eine zu geringe Anzahl kann möglicherweise das Finden eines Minimums der Verlustfunktion unterbinden.

- ▶ **Patience:** legt fest, wie viele Epochen maximal durchgeführt werden, ohne dass die Verlustfunktion weiter reduziert wird. Eine höhere patience ermöglicht, weitere Minima zu finden, impliziert dadurch jedoch eine längere Laufzeit durch mehr Epochen. Dadurch sinkt die Wahrscheinlichkeit, dass ein lokales Minimum der Verlustfunktion für eine Beendigung des Trainingsprozesses sorgt.
- ▶ **Optimizer:** steuert die Aktualisierung der weights. Mögliche optimizer sind Stochastic Gradient Descent (SGD), Momentum (Polyak 1964), Adam (Kingma et al. 2015), etc. Diese Algorithmen haben unterschiedliche Eigenschaften, allerdings gibt es keine Methode, die eindeutig zu einer besseren Performance führt als die anderen (Schaul et al. 2014).
- ▶ **Learning rate:** bestimmt, wie stark die weights zwischen den Epochen verändert werden. Eine zu geringe learning rate benötigt viele Epochen, da die Minimierung der Verlustfunktion nur langsam geschieht; eine zu hohe learning rate kann durch zu große Sprünge in den weights verhindern, dass ein Minimum gefunden wird. Bei adaptiven optimizern wie z.B. Adam ist die learning rate kein fester Wert, sondern nimmt mit zunehmenden Epochen ab (Goodfellow et al. 2016).

Die Hyperparameter des Neuronalen Netzes wurden in einer Gittersuche (engl.: grid search) optimiert. Dafür wird für jeden der zu optimierenden Hyperparameter eine Auswahl an Werten festgelegt und jede mögliche Kombination dieser Werte getestet. Die Anzahl an Hyperparameter-Kombinationen ist dabei das Produkt der Anzahl der Werte eines jeden Hyperparameters. Die Gittersuche erfolgt innerhalb einer 5-fachen Kreuzvalidierung (engl.: 5-fold [single] cross-validation), damit jeder Datenpunkt bei der Optimierung der Hyperparameter berücksichtigt werden kann. Für die 5-fache Kreuzvalidierung wird der Datensatz in 5 Teildatensätze geteilt, von denen jeweils einer als Testdaten und die anderen vier als Trainingsdaten (bzw. im Fall des NN als Trainings- und Validierungsdaten im Verhältnis 80:20) dienen. Für jeden der fünf Trainingsdatensätze wird das Modell mit jeder Hyperparameter-Kombination aus der Gittersuche trainiert und jeweils Wahrscheinlichkeiten (der Klassenzugehörigkeit) für das Testset vorausgesagt. Die Gesamtheit der fünf Testsets bildet den kompletten Datensatz ab, sodass insgesamt für jeden Datenpunkt genau eine Vorhersage erfolgt, die mit den tatsächlichen Klassen abgeglichen werden kann. Dabei wird der Entscheidungsschwellenwert (engl.: decision threshold) variiert um den besten macro-averaged F1-Score zu erhalten. Somit ergibt sich für jede Hyperparameter-Kombination ein optimierter F1-Score (mit entsprechendem Entscheidungsschwellenwert), der über den kompletten Datensatz berechnet wird und der als Kriterium dient, die beste Hyperparameter-Kombination auszuwählen. Die Einschätzung der Performance in einer Kreuzvalidierung mit Hyperparameter-Optimierung kann allerdings verfälscht sein, da die Hyperparameter jeweils für die Testsets optimiert wurden und die Vorhersage nicht auf ein unabhängiges Testset erfolgte. Um die tatsächliche Performance des Modells zu ermitteln, wurde eine nested cross-validation verwendet.

Eine *nested cross-validation* besteht aus einer *single cross-validation* (äußere Schleife), in deren Trainingssets wiederum eine *single cross-validation* durchgeführt wird (innere Schleife). In der inneren Schleife wird eine Hyperparameter-Optimierung mit Gittersuche und Optimierung des Entscheidungsschwellenwerts durchgeführt (wie oben beschrieben). Für jede Iteration der inneren Schleife (also für jedes äußere Trainingsset) ergibt sich eine Hyperparameter-Kombination mit einem Entscheidungsschwellenwert, die den höchsten F1-Score erzielt. Diese Parameter werden in der äußeren Schleife für das Trainieren des Modells auf dem jeweiligen Trainingsset bzw. für die Vorhersage der Klassen aus den Wahrscheinlichkeiten genutzt. Daraus ergibt sich für jeden äußeren *fold* ein F1-Score. Diese Scores werden zu einem finalen F1-Score

gemittelt, der die Performance des Modells am realistischsten einschätzt, da die Berechnung des Scores auf Vorhersagen auf den gesamten, ungesehenen (also nicht für die HPO verwendeten) Datensatz beruht. Um ein finales Modell für die Vorhersage von Konzentrationsklassen für neue Datenpunkte zu erhalten, wird der gesamte Datensatz mit den in der HPO ermittelten Hyperparameter-Kombination trainiert.

Die aggregierten Daten wurden mit einem *feedforward neural net* (FNN) verarbeitet. Das FNN ist eine der einfachsten Formen von neuronalen Netzen, da es die Übergabe von Signalen zwischen den Neuronen nur in eine Richtung erlaubt (Goodfellow et al. 2016). Die Implementierung des FNN erfolgte mit der Python-Bibliothek *Tensorflow* (Abadi et al. 2016).

Für die Verarbeitung der zeitlich aufgelösten Daten wurden zwei Modellkonstruktionen neuronaler Netze getestet: ein rekurrentes neuronales Netz (*engl: recurrent neural network*; RNN) und ein „hybrides“ FNN (hFNN).

Ein RNN ist ein Netzwerk, bei dem die Neuronen nicht nur in eine Richtung miteinander verknüpft sind (wie beim FNN), sondern auch rückwärtige Verknüpfungen sowie Verknüpfungen zwischen Neuronen einer Schicht bestehen. Dadurch können RNN auch zeitliche Informationen verarbeiten (Goodfellow et al. 2016) und werden häufig zur Vorhersage von Zeitreihen verwendet. Üblicherweise wird die Entwicklung einer Zielvariablen in der Zukunft, basierend auf ihrer Entwicklung in der Vergangenheit (in Abhängigkeit eines Sets an Features) vorhergesagt (Connor et al. 1994). Dabei werden pro Standort eigene zeitliche Modelle trainiert. In der Anwendung im Rahmen des FARM-Projekts soll sowohl die zeitliche Entwicklung, als auch die räumliche Verteilung berücksichtigt werden. Im Vordergrund steht hierbei weniger die Projektion von Stoffkonzentrationen in zukünftigen Jahren, als das Erlernen der Zusammenhänge zwischen Stoffkonzentrationen und räumlich-zeitlichen Feature-Variationen. Durch die Berücksichtigung der räumlichen und der zeitlichen Dimension ist die Feature-Matrix dreidimensional. Für n_f Features, n_m Messstellen und n_t Zeitpunkte hat die Feature-Matrix die Größe $n_f \times n_m \times n_t$. Hierbei gilt es zu beachten, dass die Feature-Matrix keine fehlenden Werte aufweisen darf. Daher können nur Messungen benutzt werden, für die über alle (ausgewählten) Messstellen die gleichen Features sowie die gleiche Zeitspanne an Features und Wirkstoff- bzw. Metabolitenkonzentrationen vorhanden sind. Im Fall der FARM-Datenbank führt dies zu einer erheblichen Reduzierung der verfügbaren Datenpunkte (siehe 2.2.3.5).

Alternativ zum RNN kann die zeitliche Dimension auch in einem hFNN berücksichtigt werden. Beim hFNN hat die Feature-Matrix wie üblich nur zwei Dimensionen ($n_f \times n_m$), jedoch wird der Zeitpunkt (hier: das Jahr) mit als Feature aufgenommen. Für jede Messstelle gibt es in der Feature-Matrix also so viele Zeilen, wie die Anzahl der Jahre, in denen die jeweilige Messstelle beprobt wurde. Die Anzahl der Jahre kann zwischen den Messstellen variieren. Dies birgt den Vorteil, dass alle verfügbaren Messungen (für die jeweilige Feature-Auswahl) mit einbezogen werden.

Logistische Regression

Zur binären Klassifikation eignet sich eine Form der multiplen Regression für dichotome abhängige Variablen (nur zwei diskrete mögliche Werte): die binomiale logistische Regression. Dabei wird die Wahrscheinlichkeit, dass eins von zwei möglichen Ereignissen eintritt, anhand einer S-förmigen logistischen Funktion geschätzt, die Werte zwischen 0 und 1 annehmen kann (natürlicher Logarithmus der *Odds*) (Gareth et al. 2023). Die logistische Regression stellt ein einfacheres Verfahren als Neuronale Netze oder Random-Forest dar und wurde als Benchmark-Modell herangezogen.

2.2.3.5 Feature-Auswahl und Modellvarianten

Die in der Datenbank abgelegten Features wurden hinsichtlich ihrer weiteren Verwendbarkeit für maschinelles Lernen geprüft. Die Feature-Auswahl wurde fachlich und in Bezug auf Datenqualität und -verfügbarkeit abgewogen. Die Koordinaten der Messstellen wurden nicht als Features berücksichtigt, um ein Lernen anhand eines zu spezifisch auf die einzelne Messstelle verweisenden Features zu verhindern. Andernfalls besteht die Gefahr, dass das Modell Vorhersagen vorwiegend aus einem einer Messstelle eindeutig zuzuordnenden Feature-Wert (sog. *unique identifier*) ableitet. Dementsprechend kann auch die Messstellen-Bezeichnung grundsätzlich nicht als Feature verwendet werden. Ebenso wurde das Bundesland der Messstelle nicht als Feature mitaufgenommen, da ein Bias aufgrund unterschiedlicher Messprogramme und unterschiedlicher Analytik von den eigentlich zu erforschenden naturräumlichen Zusammenhängen ablenken würde.

Zur Charakterisierung der intrinsischen Grundwasservulnerabilität wurden hydrogeologische, bodenkundliche und meteorologische Features sowie die Topographie berücksichtigt. Gleichzeitig eignen sich diese Daten, da sie für einen Großteil der Messstellen verfügbar sind und als bundesweit harmonisiertes Produkt vorliegen. Im Endeffekt ähneln die zur Bewertung der intrinsischen Grundwasservulnerabilität ausgewählten Features grundsätzlich den Variablen, die in gängigen Vulnerabilitätsindizes verwendet werden, ohne diese exakt nachbilden zu wollen. Das DRASTIC-Modell (Aller et al. 1987) und darauf aufbauende Derivate basieren im Wesentlichen auf den folgenden im Akronym enthaltenen Faktoren:

- ▶ *Depth to Water* (Grundwasserflurabstand)
- ▶ *Recharge* (Net Recharge) (Grundwasserneubildung)
- ▶ *Aquifer Media* (Material des Grundwasserleiters)
- ▶ *Soil Media* (Bodenmatrix)
- ▶ *Topography (Slope; Hangneigung)*
- ▶ *Impact of the Vadose Zone* (Einfluss der ungesättigten Zone)
- ▶ *Conductivity (Hydraulic Conductivity of the Aquifer; hydraulische Leitfähigkeit des Aquifers)*

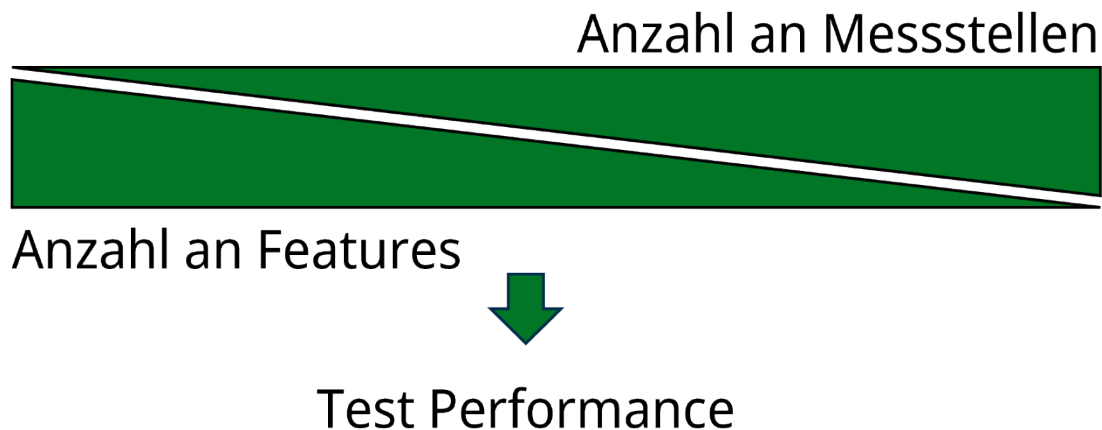
Alternativ zum Grundwasserflurabstand wurde die Tiefe der Filteroberkante verwendet und die Hohlraumart wurde als aussagekräftiges Feature anstelle des Aquifermaterials verwendet.

Die an DRASTIC angelehnten Features dienen vorrangig der Charakterisierung intrinsischer Grundwasservulnerabilität. Ein Kernanliegen dieses Projekts ist es, den Einfluss der Landnutzung in stoffspezifischen Modellen zu untersuchen. Daher wurden die genannten Features um die Anbaufläche der Kultur, für die ein Wirkstoff zugelassen ist, erweitert.

Die Datenbank umfasst zahlreiche Features, die z.T. korreliert sind. Grundsätzlich sollten stark miteinander korrelierte Features nicht gleichzeitig im maschinellen Lernen (bzw. bei statistischen Methoden insgesamt) verwendet werden, da die geschätzte Modellgüte verfälscht werden könnte und der Einfluss einzelner Features schwerer zu differenzieren ist. Bei mehreren Features aus der gleichen Domäne, die ähnliche Prozesse bzw. räumliche Gegebenheiten repräsentieren, musste daher eine Selektion stattfinden. Eine erste Reduktion der Anzahl der anfangs verfügbaren Features wurden somit anhand von linearen bivariaten Korrelationsanalysen mit allen weiteren Standort-Features vorgenommen.

Ein weiterer Grund zur gezielten Vorauswahl von Features ist die heterogene Datenverfügbarkeit. Bei Verwendung zahlreicher Features führt diese zu einer starken Reduktion der Messstellen, für welche diese Features komplett zur Verfügung stehen. Gängige Implementierungen von Random-Forest und Neuronalen Netzen bedürfen kompletter Datenreihen (ohne fehlende Werte). Die Abwägung zwischen Anzahl der Features und verfügbarer Anzahl der Messstellen (kleinster gemeinsamer Nenner aller Features) führt dazu, dass nicht alle Features Verwendung finden können (Abbildung 4).

Abbildung 4: Abwägung zwischen der Anzahl an Messstellen und Anzahl an Features mit dem Ziel einer möglichst hohen Modellgüte.



Quelle: eigene Darstellung BGR.

Des Weiteren ist darauf zu achten, dass die Anzahl der Datenpunkte ausreichend ist. Eine Hyperdimensionalität liegt vor, wenn das Verhältnis zwischen der Anzahl der Features und der Anzahl der Datenpunkte stark zu Ungunsten der Anzahl der Datenpunkte ausfällt. Dadurch kann der Lernerfolg der Modelle behindert werden und es wird Overfitting riskiert.

Grundsätzlich sollten Features auch mit der jeweiligen Zielvariable korreliert sein, daher wurde zuvor eine Korrelationsanalyse durchgeführt. Diese eignet sich gut als erste Einstufung. Lineare bivariate Korrelationen zwischen einem einzigen Feature und der Zielvariable spiegeln den Einfluss des Features jedoch oft unvollständig wider, da Interaktionen mit anderen Features und nichtlineare Zusammenhänge nicht berücksichtigt werden. Es wurden sich an signifikanten Korrelationskoeffizienten zwischen den Features und der Zielvariable orientiert.

In einem weiteren Schritt wurde die Relevanz einzelner Features auch anhand ihrer Feature Importance bewertet. Es wurden iterativ mehrere Features-Auswahlen untersucht, indem Modellläufe mit unterschiedlichen Features durchgeführt und die Features aussortiert wurden, die keinen oder einen sehr geringen Einfluss auf die Vorhersage der Zielvariable aufwiesen.

Eine vollständige Eliminierung von Korrelationen zwischen Features ist faktisch nicht zu erreichen. Sehr starke Korrelationen sollten aufgrund von Redundanz vermieden werden, weniger starke Korrelationen ergeben sich allerdings aus der Natur der Daten und der räumlichen Zusammenhänge. Zahlreiche Features zeigen Korrelationen untereinander auf, da sie gleichermaßen als Zeiger für landwirtschaftliche Aktivität dienen. Weiterhin ist davon auszugehen, dass die räumliche Verteilung von Ackerland in Deutschland an naturräumlichen Begebenheiten, Boden- und Klimaräume sowie Höhenlagen gebunden ist. Variablen wie die Nitratkonzentration im Grundwasser können als räumliche Proxys verstanden werden, aber auch als eine Art „Co-Zielvariable“, da neben Pflanzenschutzmitteln auch Nitrat einen Hinweis

auf das Versickerungsverhalten und den weiteren Transport von Stoffen an einem Standort liefern kann. Die Features wurden daher hinsichtlich der Interpretierbarkeit in drei Gruppen unterteilt.

Gruppe a) stellt eine reduzierte Anzahl an Features da, die sich auf möglichst unkorrelierte Features und sich damit auf das Erlangen von Prozessverständnis konzentriert. Da die Sand-, Ton- und Schluffgehalte des Bodens sich gegenseitig vollumfassend erklären und unter diesen Features auch starke bivariate lineare Korrelationen vorliegen, wurde für Gruppe a) ausschließlich der Sandgehalt gewählt. Die Makroporenklasse ergibt sich vorwiegend aus den Ton- und Schluffgehalten. Somit wurden auch diese Bodentextureigenschaften mitberücksichtigt, wobei die Makroporenklasse deutlich weniger mit dem Sandgehalt korreliert als Schluff und Ton für sich. Die Grundwasserneubildung vereint Aspekte der Grundwasserleiterüberdeckung mit langjährigen klimatologischen/meteorologischen Bedingungen (Anhang A.2).

a) Reduzierte Feature-Auswahl

- ▶ Flächen im Puffer einer Messstelle pro Kulturklasse, für die der jeweilige Wirkstoff zugelassen ist
- ▶ Tiefe der Filteroberkante unter Geländeoberkante
- ▶ Sandgehalt im Boden
- ▶ Organischer Kohlenstoffanteil im Boden
- ▶ Makroporenklasse
- ▶ Langjährige Rate der Grundwasserneubildung
- ▶ Hohlraumart des Aquifers
- ▶ Hydraulische Leitfähigkeitsklasse des Aquifers

Weitere Features sind der hydrogeologische Großraum, die Lufttemperatur, der Niederschlag, die Geländehöhe, die Hangneigung.

Gruppe b) umfasst somit die Features aus Gruppe a) sowie folgende weitere Features:

b) Erweiterte Feature-Auswahl

- ▶ Hydrogeologischer Großraum
- ▶ langjähriges Mittel der Lufttemperatur
- ▶ mittlerer Sommerniederschlag
- ▶ mittlerer Winterniederschlag
- ▶ Verhältnis aus Niederschlag April-September/Okttober-März
- ▶ Geländehöhe
- ▶ Hangneigung
- ▶ Gesteinsart
- ▶ Geochemischer Typ

- ▶ DSD_1, DSD_2, LP_1, LP_2, SD_1, SD_2: Aus MOHP, abstrahierte Abstände zu Einzugsgebietsgrenzen und Gewässer für Strahler-Ordnung 1 und 2
- ▶ Schluffgehalt im Boden
- ▶ Tongehalte im Boden
- ▶ Ackerfläche im Puffer (kulturunspezifisch)
- ▶ Sickerwasserrate

Features, die aus den grundwasserchemischen Messwerten an den Messstellen abgeleitet wurden, und als zusätzliche Zeiger von Grundwasserqualität gesehen werden können, sind nachfolgend gelistet. Gruppe c) umfasst somit alle Features aus Gruppen a) und b) als auch:

c) Zusätzliche Qualitätszeiger/ „Co-Zielvariablen“

- ▶ mittlere Nitratkonzentration an der Messstelle
- ▶ mittlere Ammoniumkonzentration
- ▶ mittlere Magnesiumkonzentration
- ▶ mittlere Sulfatkonzentration
- ▶ mittlere Chloridkonzentration
- ▶ mittlere Kaliumkonzentration
- ▶ mittlere Natriumkonzentration

Aus diesen drei Feature-Gruppen wurden folgende Modellvarianten entwickelt. Modellvariante 1) umfasst ausschließlich die Features aus Gruppe a). Modellvariante 2) umfasst alle genannten Features (Gruppen a), b) und c)).

Der Modellvergleich lässt sich folglich zusammenfassen als:

1) *Prozessorientierte Features (nicht korreliert)*

2) *Prozessorientierte (auch korrelierte) + räumliche Features + GW-Chemie*

Ziel des Modellvergleichs ist es, abzuschätzen, inwieweit die Modellgüte von der Anzahl an berücksichtigten Features abhängt. Zusätzlich soll geprüft werden, ob eine solche Trennung der Standortfaktoren zur Interpretation der Modelle beiträgt. Es soll versucht werden, Standortfaktoren, von denen angenommen wird, dass sie zum Prozessverständnis beitragen, von ausschließlich räumlich bedingten Korrelationen möglichst zu trennen. Es wird angenommen, dass ein Großteil der zur Verfügung stehenden Features sowohl Informationen zum Prozess (stoffspezifische Grundwasservulnerabilität an einem Standort), als auch räumliche Zusammenhänge in sich vereinen.

2.2.3.6 Stoffauswahl für das Maschinelle Lernen und Konzentrationsklassen

Die Auswahl geeigneter Stoffe sowie die Einteilung in Konzentrationsklassen für das Maschinelle Lernen orientierte sich an der in 2.2.1 beschriebenen Problematik des Anteils zensierter Daten und der Gewährleistung einer ausreichenden Anzahl an Messstellen und damit verbundener, ausreichender Modellgüte (siehe 2.2.3.3).

Zur Einteilung der Konzentrationen von Wirkstoffen und Metaboliten in sinnvolle Konzentrationsbereiche, die auch zwischen Konzentrationen unterhalb der $0,1 \mu\text{g L}^{-1}$ Grenze differenzieren, wurden zensierte Daten mit geschätzten Werten aufgefüllt. Dafür wurden mittels MLE die Parameter einer logarithmischen Normalverteilung (Mittelwert und Standardabweichung) anhand der Messwerte aller Messstellen geschätzt. Basierend auf diesen Parametern wurden mögliche Konzentrationen generiert und mögliche Werte unterhalb der jeweiligen LOQ zufällig gezogen. Messwerte oberhalb der LOQ wurden beibehalten. Somit ergibt sich ein zusammengesetzter Datensatz, bei dem zensierte Werte geschätzt und aufgefüllt wurden. Der Prozess des Auffüllens betrifft ausschließlich zensierte Werte mit einer LOQ, die über $0,05 \mu\text{g L}^{-1}$ liegt. Eine geringe Anzahl an Messungen weist eine LOQ von $0,1 \mu\text{g L}^{-1}$ auf. Bei den in den Modellen verwendeten Stoffen und im Zeitraum ab 2010 betrifft dies nur 1,96% der Messungen. Der Großteil der zensierten Daten ist hiervon nicht betroffen, da die meisten LOQ darunterliegen, und fällt somit in die Klasse 1 ($< 0,05 \mu\text{g L}^{-1}$). Die zufällige Zuordnung soll dazu dienen, den geschätzten Wert nicht direkt von der Höhe der LOQ abhängig zu machen. Hierdurch ergibt sich eine gewisse Unsicherheit. Dieser Fehler wurde allerdings als gering betrachtet, auch in Hinblick darauf, dass wenige Messungen von einer hohen LOQ betroffen sind und auch zeitlich aggregierte Konzentrationsklassen in den Modellen Verwendung fanden. Eine zweite Klasse umfasst Werte von $0,05$ bzw. größer als $0,05$ und unterhalb von $0,5 \mu\text{g L}^{-1}$. Die dritte Klasse bezieht sich dementsprechend auf Konzentrationen oberhalb von $0,5 \mu\text{g L}^{-1}$. Anstatt eines 3-Klassen-Modells wurde auch eine binäre Klasseneinteilung getestet, wobei Klassen 2 und 3 zusammenfallen.

Die gewählten Klassengrenzen orientieren sich nicht an regulatorischen Grenzwerten, sondern sind der Datenverteilung geschuldet und dienen der Erzeugung von einer möglichst ausgeglichenen Anzahl an Messstellen pro Klasse. Nach wie vor fällt ein Großteil der Daten in Klasse 1, dennoch wurde mit der genannten Klasseneinteilung ein Kompromiss gewählt, der die Vorhersagbarkeit von Klassen 2 und 3 gewährleisten soll.

Als alternative Zielvariable wurden auch Perzentile einer Kaplan-Maier-Verteilung berechnet und ihrer Rangordnung nach durchnummeriert. Diese „Ranks“ wurden in Regressionsanalysen getestet. Allerdings erwies sich der hohe Zensurgrad für einen Großteil der betrachteten Stoffe als problematisch, sodass die erzeugten Kaplan-Maier-ranks gerade im niedrigen Konzentrationsbereich eher für Verzerrungen sorgten. Außerdem wurde ein Durchpausen der Struktur der LOQ beobachtet, wodurch die auf den Kaplan-Maier-Perzentilen beruhenden Ranks auf bestimmte Regionen und Bundesländer hinweisen und somit Artefakte erzeugen. Daher wurde dieser Ansatz nicht weiterverfolgt.

Die Stoffauswahl wurde iterativ immer weiter reduziert, bis sich auf acht Metaboliten mit ausreichender Datengrundlage verständigt wurde. Diese sind:

- ▶ Metolachlor ESA (persistent, sehr mobil)
- ▶ Metolachlor OXA (persistent, sehr mobil)
- ▶ Metolachlor Metabolit NOA 413173 (persistent, sehr mobil)
- ▶ Metazachlor ESA (persistent, sehr mobil)
- ▶ Metazachlorsäure (moderat persistent, sehr mobil)
- ▶ Desphenyl-Chloridazon (persistent, mobil)
- ▶ Methyl-desphenyl-Chloridazon (persistent, mobil)

- ▶ Trifluoressigsäure (TFA) (sehr persistent, sehr mobil)

Anhand der beschriebenen Konzentrationsklassengrenzen wurde für Metaboliten die binäre (2 Klassen; Klassengrenze $0,05 \mu\text{g L}^{-1}$) sowie Mehrklassen-Klassifikation (3 Klassen; dritte Klasse $> 0,5 \mu\text{g L}^{-1}$) gewählt.

Bei den Wirkstoffen wurden binäre Stoffgruppen- und Einzelstoffmodelle mit der Klassengrenze $0,01 \mu\text{g L}^{-1}$ gerechnet (bezogen auf Gesamtzeitraum). Einzelstoffmodelle für Wirkstoffe wurden für Bentazon, Atrazin und Simazin trainiert. Wirkstoffe wurden anhand ihrer Mobilität (beschrieben in 2.1.2) in drei Gruppen eingeteilt, wobei mobile und sehr mobile Wirkstoffe zusammengenommen wurden. Wurde diese Klassengrenze stoffunabhängig im Gesamtzeitraum an einer Messstelle überschritten, wird diese Messstelle dieser Klasse zugeordnet. Somit wurden Wirkstoffgruppenmodelle für

- ▶ sehr mobile und mobile,
- ▶ moderat mobile,
- ▶ wenig mobile Wirkstoffe

gerechnet.

Zusammenfassend sind für die nachfolgenden Analysen die folgenden Zielvariablen gewählt worden:

- ▶ Metaboliten: Einteilung in Konzentrationsklassen 1) $< 0,05$, 2) $\geq 0,05$ und $< 0,5$, sowie 3) $> 0,5 \mu\text{g L}^{-1}$. Für eine Mehrheit der Messungen ist eine eindeutige Einteilung möglich, für einen geringen Anteil wurde eine Schätzung der Verteilungsparameter und darauf basierend eine Auffüllung zensierter Werte vorgenommen.
- ▶ Für Analysen, die Wirkstoffe betreffen, wurde nicht nur die Grenze von $0,1 \mu\text{g L}^{-1}$, sondern auch $0,01 \mu\text{g L}^{-1}$ verwendet. Da das Schätzen von Verteilungsparametern bei stark zensierten Wirkstoffen nicht vertretbare Unsicherheiten beinhalten kann, wurden keine zensierten Werte imputiert. Hierfür werden Messdaten mit höherer Bestimmungsgrenze ignoriert, was zu einem Verlust an Messstellen führen kann.

2.2.3.7 Erklärbarkeit und Interpretierbarkeit von Modellen

Um die Wichtigkeit einzelner Features für die Modellvorhersage zu untersuchen, kommen unterschiedliche Methoden zur Bestimmung der *Feature Importance* (Featuregewichtung) zur Anwendung. Dabei ist zum einen die Sortierung der Features gemäß ihrer Bedeutung für den Modelloutput von Interesse (Ranking). Zum anderen werden Methoden verwendet, die auf die *Feature Effects* (Featureeffekt) abzielen, um auch die konkrete Wirkung des Features auf die vorhergesagte Zielvariable in verschiedenen Wertebereichen zu untersuchen. Interaktionen zwischen Features und deren kombinierten Effekten auf die Zielvariable können durch kombinierte Plots der Feature Effects dargestellt werden. Das Feature Importance Ranking ergibt sich aus der Gesamtheit und der Relevanz der Effekte eines Features auf die Vorhersage und ist durch Korrelationen und Interaktionen zwischen Features beeinflusst.

Unabhängig von der Modellwahl (model-agnostic) kann die *Permutation Importance* (Fisher et al. 2019) verwendet werden, bei der der Verlust der Vorhersagegüte nach der zufälligen Permutation (Durchmischung) eines Features berechnet wird. Starker Performanceverlust verweist im Umkehrschluss auf die Wichtigkeit des Features. Um die Permutation Importance robuster zu machen, wurde die Permutation zehnfach mit unterschiedlichen Seeds durchgeführt und der Performanceverlust gemittelt.

Als informationstheoretisches Maß der Feature Importance lässt sich beim Random-Forest die *Gini impurity* verwenden. Sie ist ein Maß dafür, wie gut eine Aufteilung (Split) die Daten in zwei homogene Gruppen trennt und kommt in Entscheidungsbaumverfahren (Random-Forest) als essenzielles Kriterium während des Lernprozesses zum Zuge (Breiman 2001). Pro Feature wird die Gini impurity über alle Bäume, in denen dieses Feature Verwendung findet, aufsummiert und normalisiert, womit Features nach ihrer relativen Bedeutung geordnet werden können. Die Gini impurity ist eine modellspezifische Methode der Feature Importance und wurde daher nur für Modellergebnisse aus Random-Forest und nicht für Neuronale Netze berechnet.

Partial dependence plots (PDP) veranschaulichen den marginalen Effekt eines Features auf die Vorhersage. Dieser Effekt kann linear, monoton oder komplex sein. In unserem Fall zeigen die PDP die Wahrscheinlichkeit einer Konzentrationsklasse für verschiedene Werte des Features. Aus den PDP kann indirekt auch die Feature Importance abgeleitet werden, wobei eine flache Kurve beispielsweise für einen relativ kleinen marginalen Effekt steht. Allerdings ist Vorsicht geboten, da potenzielle Feature Interaktionen keine Beachtung finden. Grundsätzlich handelt es sich bei PDP um eine globale Methode, bei der der durchschnittliche Effekt über alle Daten bestimmt wird (Molnar 2022).

Alle genannten Verfahren zur Bestimmung der Feature Importance weisen Vor- und Nachteile auf. So ist die Gini impurity dafür anfällig, die Bedeutung von kategorischen Features mit einer hohen Anzahl an unterschiedlichen Ausprägungen zu überschätzen. Bei der Permutation Importance kann es zu einer Über- bzw. Unterschätzung der Bedeutung von miteinander korrelierten Features kommen. Die Durchmischung der Daten bricht die Korrelation zwischen Feature und Zielvariable. Dies kann auch eventuell vorhandene Interaktionen zwischen einem Feature und weiteren Features und wiederum deren Einfluss auf die Zielvariable reduzieren. Gerade bei stark korrelierten Features sind die relativen Listenplätze im Ranking mit Vorsicht zu betrachten, da die Reihenfolge in welcher eins der beiden korrelierten Feature im Trainingsprozess eingebunden wird, auf Zufälligkeit beruht (Molnar 2022).

Die Stärke der Interaktionen zwischen zwei oder mehreren Features, die auf die Vorhersage einwirkt, beschreibt die Variation der Vorhersage bei Veränderung mehrerer Features und errechnet sich nach Dekomposition der Vorhersage in einen konstanten Term, sowie in jeweils einen Term für das erste und nachfolgende Features mittels der H-Statistik. Neben der Interaktion zwischen zwei einzelnen Features, kann die Interaktion eines Features mit allen anderen Features als übergreifende Interaktionsstärke aufgeführt werden.

3 Ergebnisse

3.1 Auswertung der Datenbank

3.1.1 Messprogramme

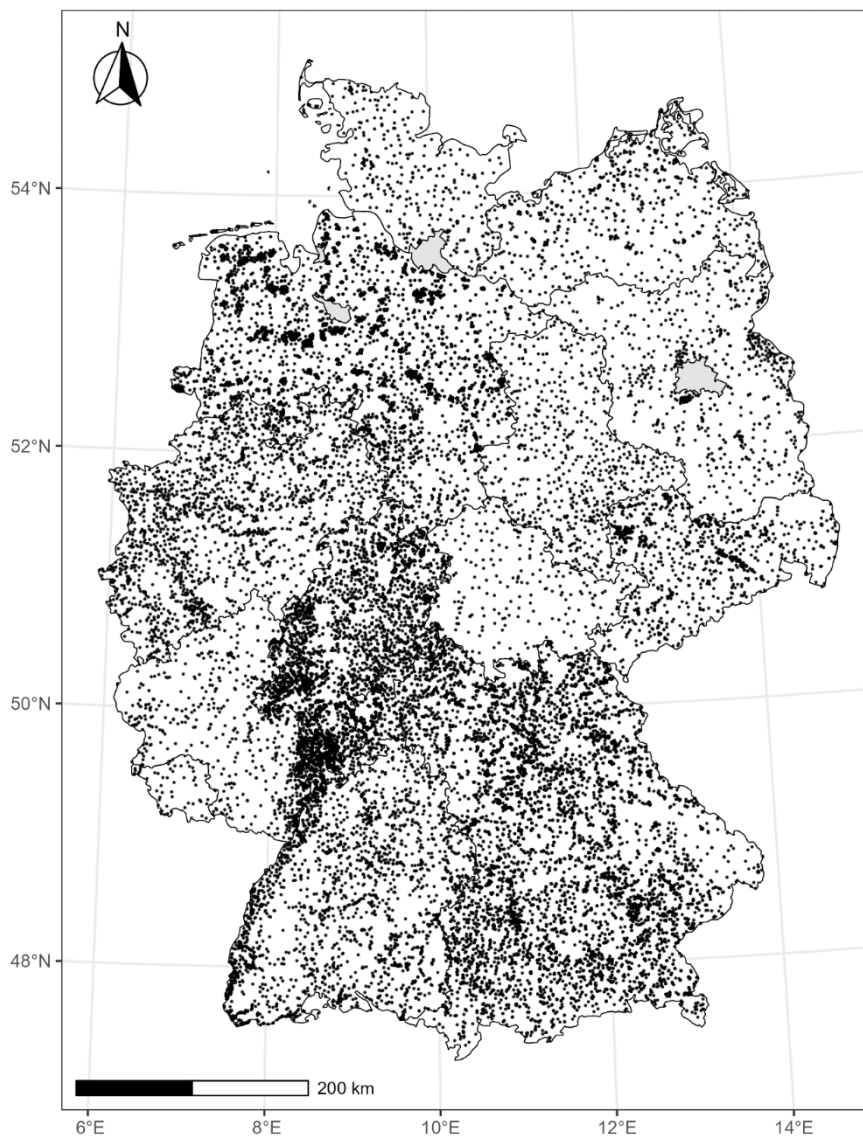
Dem Projekt wurden insgesamt 26.192 Messstellen bereitgestellt, die zwischen 1973 und 2021 auf (in Summe) 530 Wirkstoffe und Metaboliten beprobt wurden. Die räumliche Verteilung aller im Projekt vorliegenden Messstellen, an denen Messungen der Konzentrationen von Wirkstoffen bzw. Metaboliten durchgeführt wurden, wird in Abbildung 5 dargestellt. Es handelt sich zu einem Großteil um Messstellen der Länder. In Bayern, Mecklenburg-Vorpommern, Niedersachsen und Nordrhein-Westfalen wurden zusätzlich Daten der Wasserversorger mitgeliefert, welche in Bayern und Niedersachsen einige tausende Messstellen ausmachen (Abbildung 6). Die Anzahl der pro Jahr auf Wirkstoffe und Metaboliten beprobten Messstellen steigt für einen Großteil der Bundesländer mehr oder weniger stark mit der Zeit an, unterliegt jedoch auch Schwankungen (Abbildung 7). Während in Brandenburg, Mecklenburg-Vorpommern, Nordrhein-Westfalen und Sachsen die Anzahl an Messstellen kontinuierlich steigt, gibt es in Baden-Württemberg, Hessen und Niedersachsen sprunghafte Anstiege. In Rheinland-Pfalz, dem Saarland und Sachsen-Anhalt bleibt die Anzahl eher konstant. Die großen Flächenländer Baden-Württemberg, Bayern und Niedersachsen sowie Hessen stellen die meisten Messstellen. Die Gesamtverfügbarkeit der Messungen von Stoffkonzentrationen ist stark stoffspezifisch. Dabei variieren die Anzahl der beprobten Messstellen, die Zeitpunkte der ersten Messung und der Anstieg der Messungen pro Jahr. Die ersten Messungen von Stoffkonzentrationen bestimmter Stoffe im Datensatz (bspw. Bentazon und Desphenyl-Chloridazon) fanden 1973 statt. Dabei handelt es sich um Einzelmessungen, die in den Folgejahren häufig nicht wiederholt wurden. Ein leichter Zuwachs an Messstellen ist stoffübergreifend etwa ab 2007 zu verzeichnen; zum Jahr 2010 zeigen viele Stoffe einen sprunghaften Zuwachs und eine unregelmäßig steigende Tendenz in den folgenden zehn Jahren. Starke Unterschiede in der Verfügbarkeit von PSM-Messungen gibt es nicht nur zwischen den Stoffen, sondern auch für denselben Stoff zwischen verschiedenen Bundesländern. Neben der unterschiedlichen Anzahl bzw. Dichte an generell verfügbaren Messstellen zeichnen sich unterschiedliche Messnetze ab. Einige Länder verfügen ab 2010 über recht konstante Anzahlen an Messstellen; andere zeigen mehrjährige Lücken oder steigen erst später in die Messreihe ein. Die Anzahl der im Monitoring berücksichtigten Stoffe steigt über die Jahre; ein nennenswerter Anstieg ist insbesondere ab 2007 zu verzeichnen (Abbildung 11). In dieser Zeit rücken auch vermehrt Metaboliten in den Fokus der Untersuchungen (Abbildung 12).

Die Konsistenz der Messungen pro Messstelle hängt ebenfalls vom Messnetz ab. Die Messstellen werden in unterschiedlichen Messintervallen beprobt. Eine konstante Anzahl an Messungen eines bestimmten Stoffes über mehrere Jahre ist nicht mit konsistenten Messreihen an jeder Messstelle gleichzusetzen. Ein erheblicher Anteil an Messstellen wurde nur einmalig beprobt; andere mehrfach, aber mit unregelmäßigen, mehrjährigen Lücken zwischen den Messungen. Durchschnittlich wird ein Viertel aller Messungen mindestens jährlich wiederholt, Dreiviertel der Messungen werden alle 3 bis 4 Jahre oder öfter wiederholt (Abbildung 8). Im Zeitraum 2010-2019 kommen mehr als 3.000 Messstellen vor, die nur ein einziges Mal beprobt wurden, die Mehrzahl der Messstellen wurde in diesem Zeitraum jedoch häufiger beprobt (Abbildung 9). Nur ein geringer Anteil der Messstellen weist eine konstante mehrjährige Messreihe auf (Abbildung 10a). Neben den Unterschieden zwischen den Stoffen zeichnen sich auch hier deutlich die Ausrichtungen der Messprogramme der Bundesländer durch (Abbildung 10b): Während es in Niedersachsen und Baden-Württemberg eine hohe Anzahl an Messstellen bei

einer geringen durchschnittlichen Serienlänge gibt, verfügen beispielsweise Sachsen, Hessen und Brandenburg über weniger Messstellen, die dafür aber regelmäßig beprobt wurden.

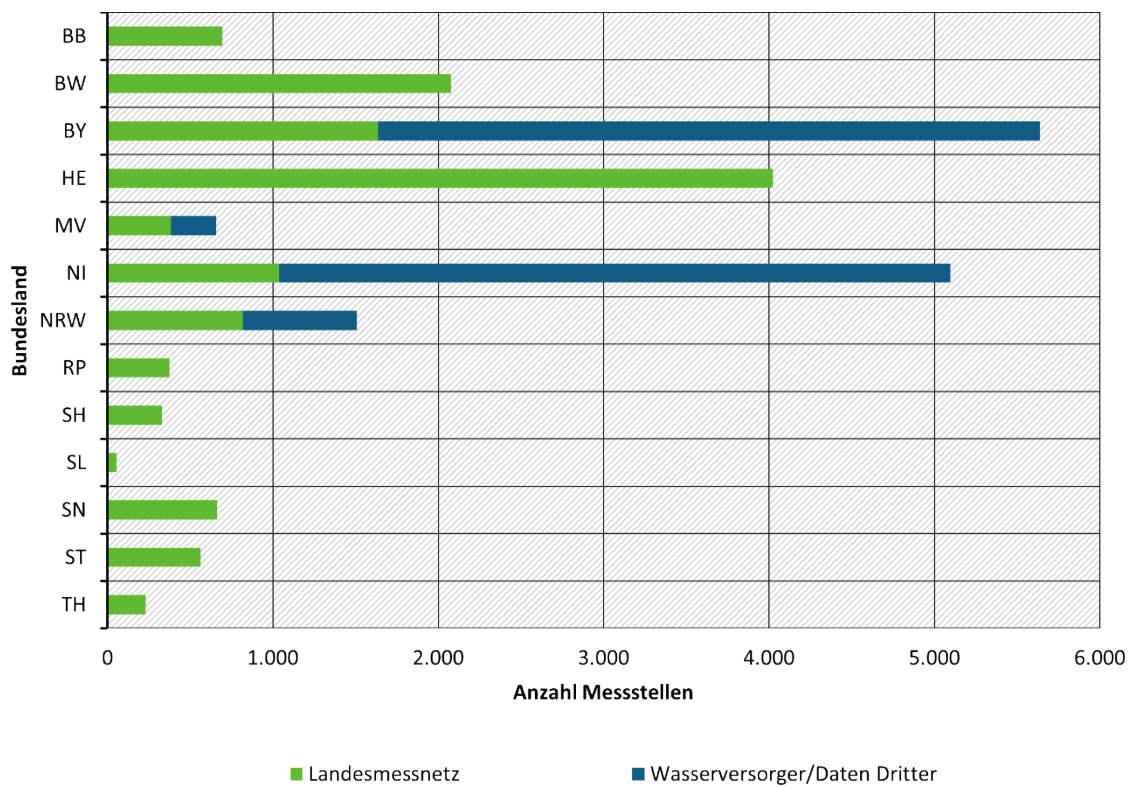
Die Messprogramme der Länder unterscheiden sich außerdem im Stoffspektrum (Abbildung 13 und Abbildung 14) und über die Zeit. Es ist festzustellen, dass in der Summe mit der Zeit deutlich mehr Stoffe in das Monitoring aufgenommen wurden. Je nach Bundesland bleibt die Stoffanzahl relativ konstant, wächst mit der Zeit an oder fluktuiert. Der Anstieg der Stoffanzahl in manchen Jahren spricht dafür, dass auch zeitlich begrenzte Sondermessprogramme in den Daten enthalten sind. Generell lässt sich sagen, dass der Datensatz ab etwa 2010 in den meisten Bundesländern eine recht umfängliche Anzahl an Messstellen und Stoffen in den Messprogrammen berücksichtigt werden. Aus diesem Grund beziehen sich ein Großteil der nachfolgenden Analysen auf diesen Zeitraum.

Abbildung 5: Gesamtheit der Messstellen im Datensatz, an denen jemals Messungen von Wirkstoff- bzw. Metabolitenkonzentrationen im Grundwasser durchgeführt wurden.



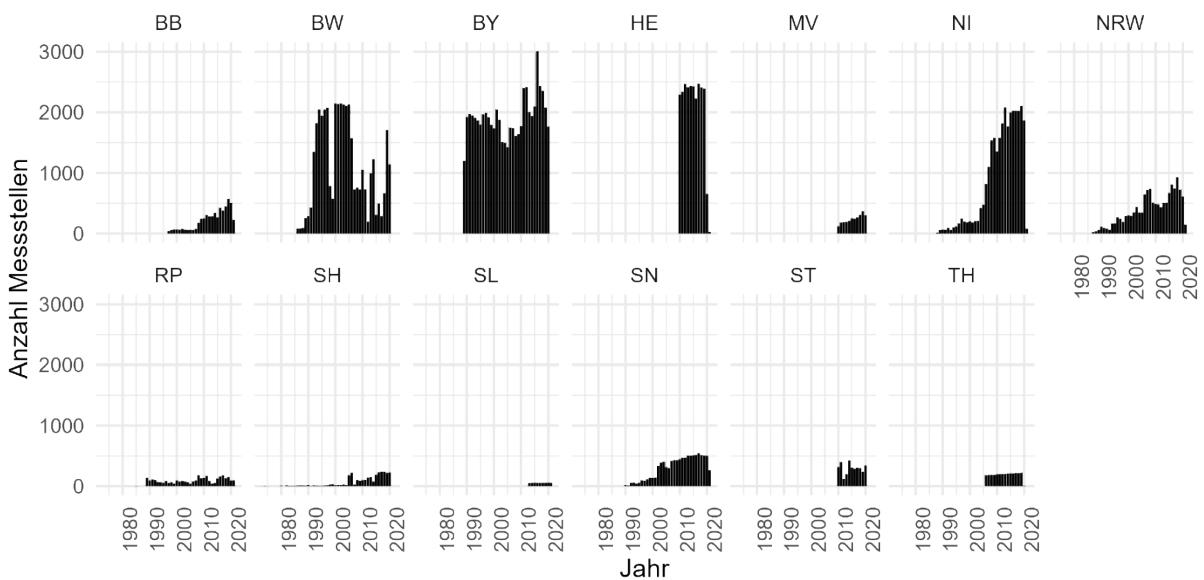
Quelle: eigene Darstellung BGR.

Abbildung 6: Anzahl der Messstellen, die in der FARM Datenbank berücksichtigt werden differenziert in Landesmessnetz und Wasserversorgern bzw. Dritten.



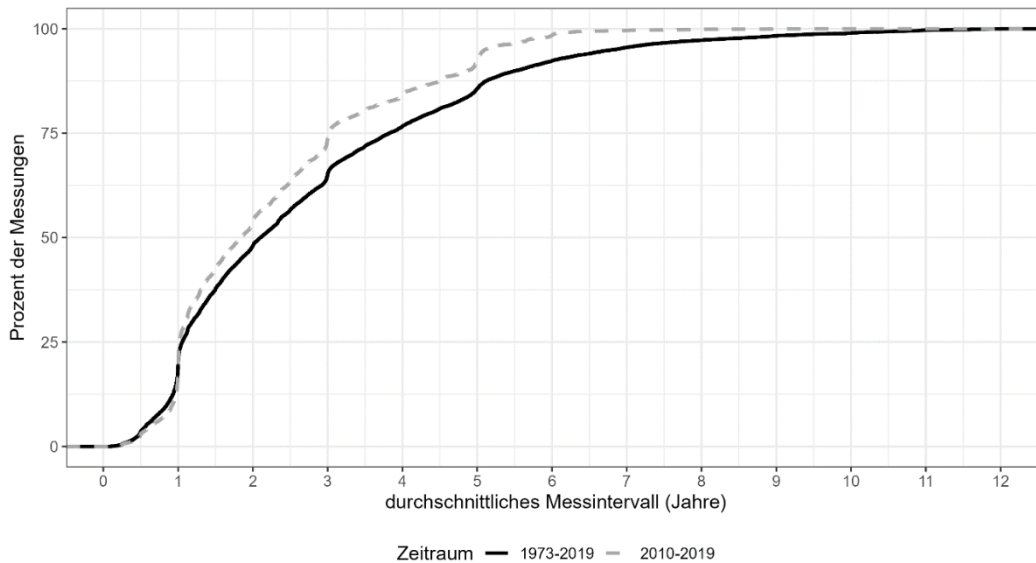
Quelle: eigene Darstellung BGR.

Abbildung 7: Anzahl der Messstellen pro Bundesland über die Zeit.



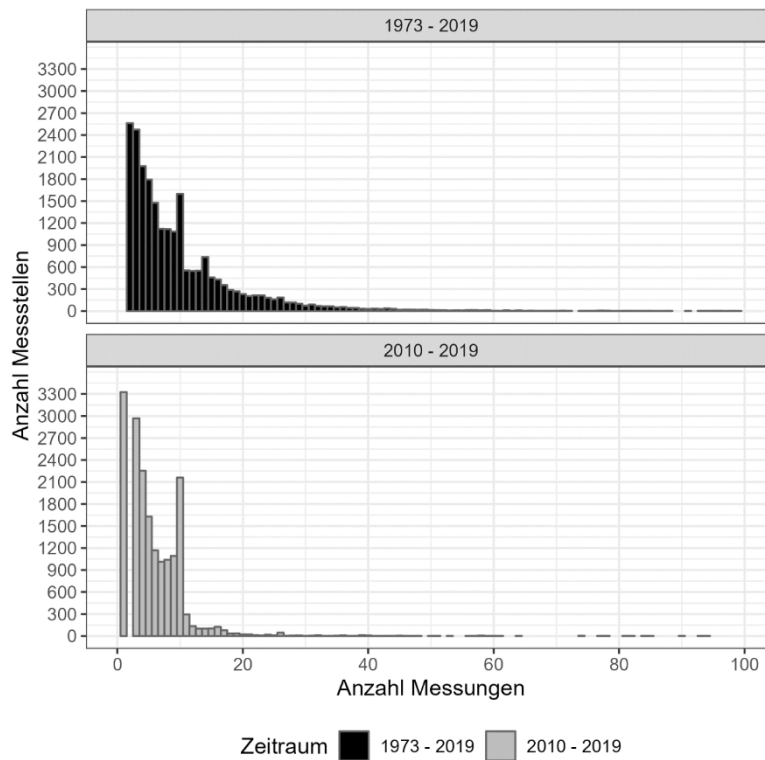
Quelle: eigene Darstellung BGR.

Abbildung 8: Das durchschnittliche Messintervall in Jahren ist gegen den prozentualen Anteil der Messungen in Form einer kumulativen Verteilungskurve aufgetragen und für den Gesamtzeitraum (1973-2019; schwarze durchgezogene Linie) bzw. für das letzte Jahrzehnt (2010-2019; graue gestrichelte Linie) differenziert dargestellt.



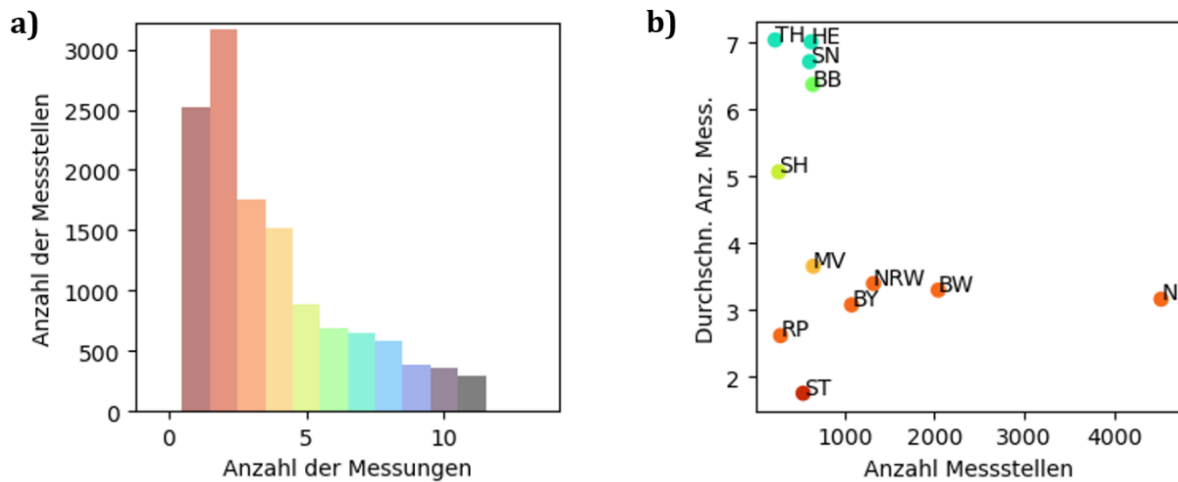
Quelle: eigene Darstellung BGR, adaptiert nach (Cooke et al. 2024).

Abbildung 9: Anzahl der Messstellen und Anzahl der insgesamt durchgeführten Messungen; jeweils dargestellt für den Gesamtzeitraum (1973-2019; oben) und für den Zeitraum von 2010-2019 (unten).



Quelle: eigene Darstellung BGR.

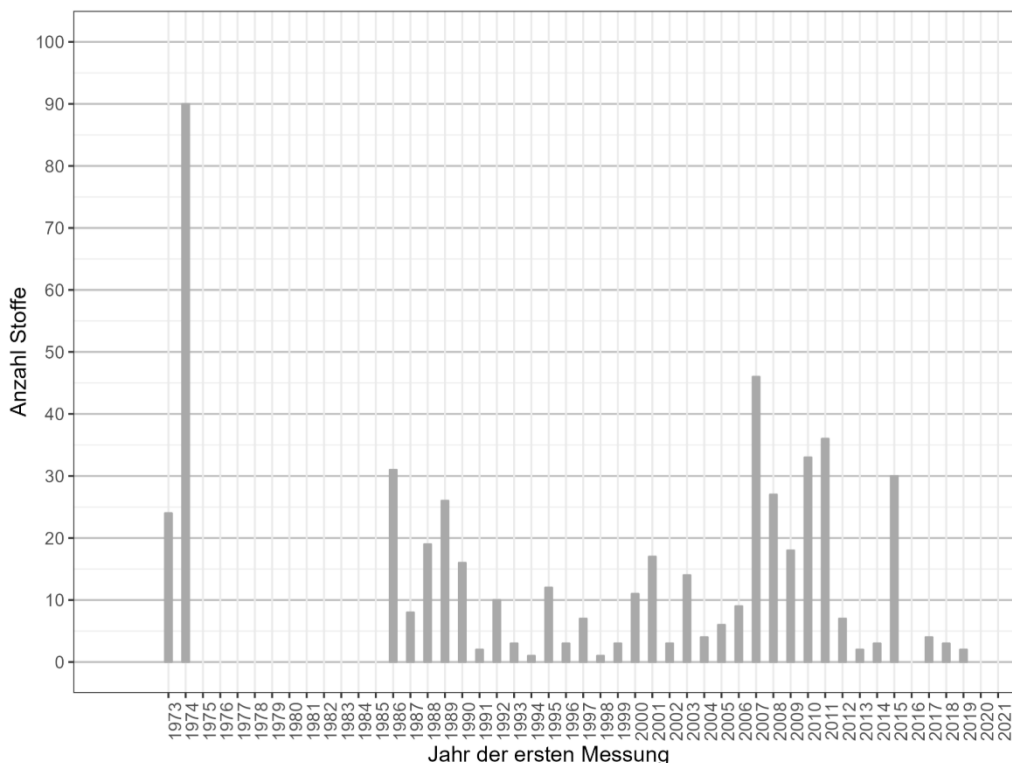
Abbildung 10: Anzahl der Messstellen pro Anzahl der Messungen für den Zeitraum 2010-2019 (a) und durchschnittliche Anzahl an Messungen pro Messstellen über Anzahl der Messungen nach Bundesland gruppiert (b) (exemplarisch für Methyldephenyl-Chloridazon).



Quelle: eigene Darstellung BGR.

Etwa 2.500 Messstellen wurden nur einfach beprobt. Die meisten Messstellen wurden zweimal beprobt. Mit zunehmender Anzahl der Messungen sinkt die Anzahl der Messstellen drastisch ab.

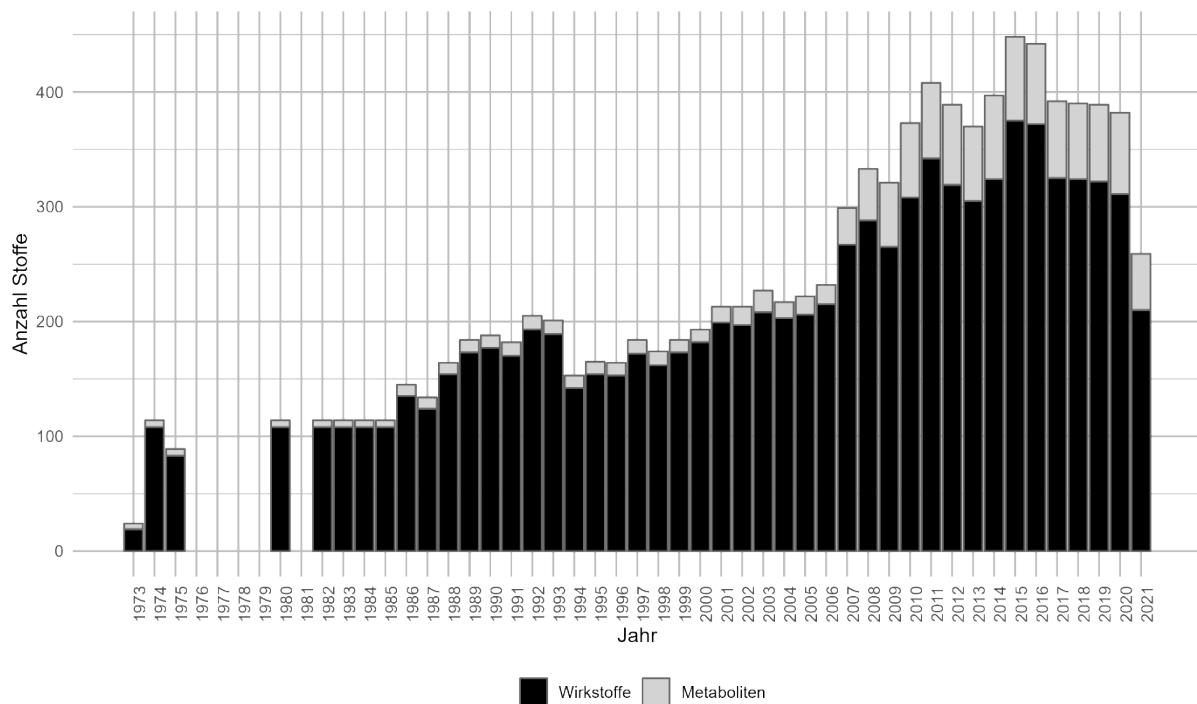
Abbildung 11: Zeitlicher Verlauf der Anzahl der Stoffe im Monitoring Programm pro Jahr der ersten Messung.



Quelle: eigene Darstellung BGR.

Mit Beginn des Monitorings 1973 kamen die meisten Stoffe hinzu. Nach einer gut zehnjährigen Pause kommen ab 1986 weitere Stoffe hinzu. Die Anzahl fluktuiert von wenigen bis über 40 Stoffen pro Jahr.

Abbildung 12: Anzahl der im Monitoring untersuchten Wirkstoffe und Metaboliten pro Jahr.



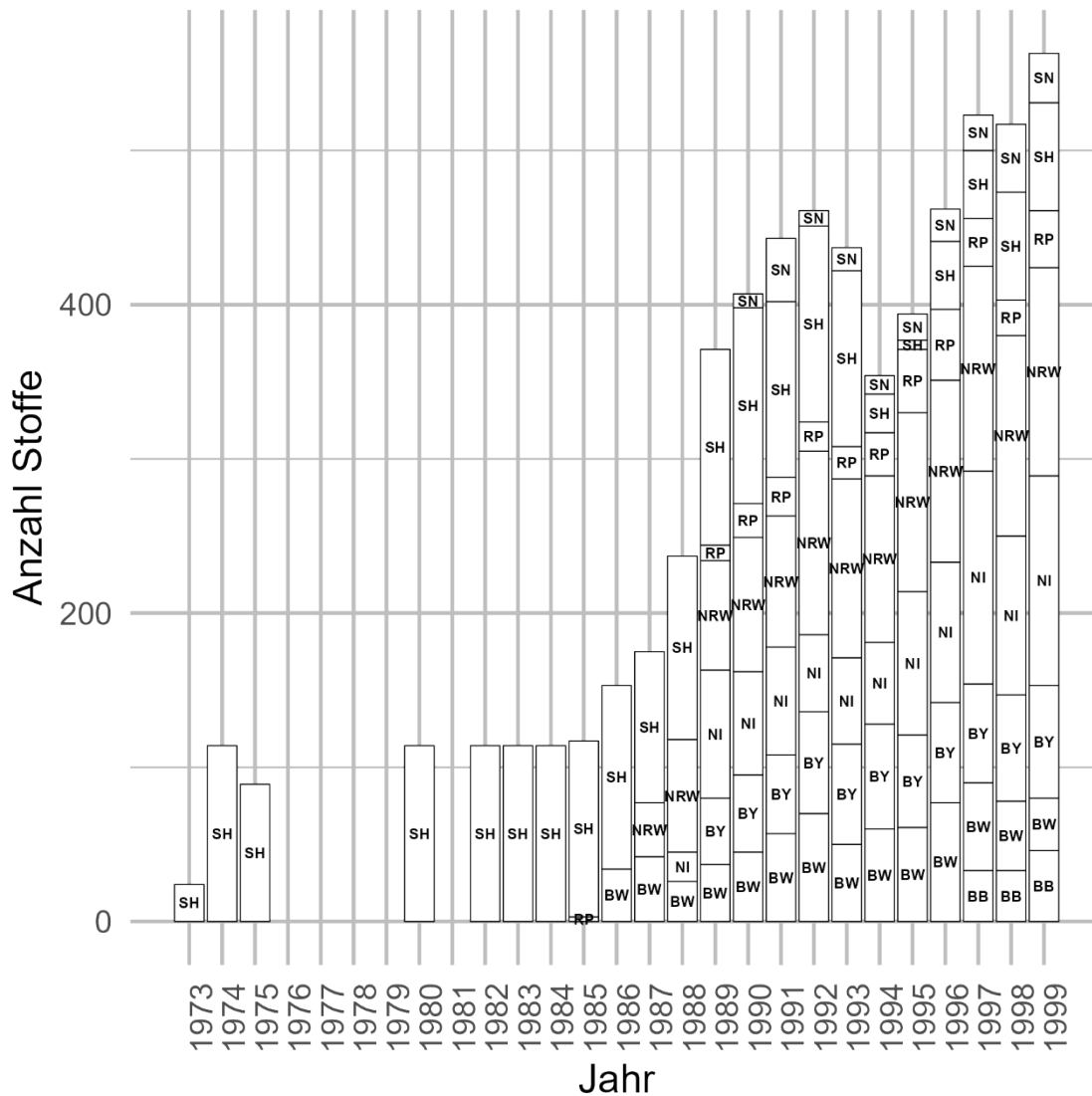
Quelle: eigene Darstellung BGR.

Ab 2007 ist ein Zuwachs der untersuchten Metaboliten zu verzeichnen. Die Wirkstoffe machen in jedem Jahr den Großteil (mindestens 80 bis über 90%) der untersuchten Stoffe aus.

3.1.2 Stoffspektrum

Im Folgenden soll das Stoffspektrum nach verschiedenen Gesichtspunkten untersucht werden. Es soll nach Bundesland, Pestizidklasse, Stoffeigenschaften und Entwicklung über die Zeit analysiert werden. Weiterhin werden die Unterschiede und Überschneidungen der beprobten Stoffe zwischen den Bundesländern ausgewertet. In einem weiteren Schritt erfolgt die Charakterisierung der Stoffe nach dem Anteil der Messungen unterhalb der Bestimmungsgrenze (Zensurgrad). Zum besseren Vergleich der Gesamtzahl an Stoffen zwischen Ländern wurde in Abbildung 15 auch nach Wirkstoffen und relevanten Metaboliten bzw. nicht-relevanten Metaboliten unterschieden (Relevanzstatus Stand März 2023). Aus Abbildung 6 und Abbildung 15 ergeben sich unterschiedliche Kombinationen aus Stoffspektrum und Messstellenanzahl in den Bundesländern, wie z.B. Hessen, das mit einer relativ großen Anzahl von Messstellen ein vergleichsweise enges Stoffspektrum beprobt. Grundsätzlich wird in allen Bundesländern größtenteils nach Wirkstoffen und relevanten Metaboliten gesucht (Abbildung 15).

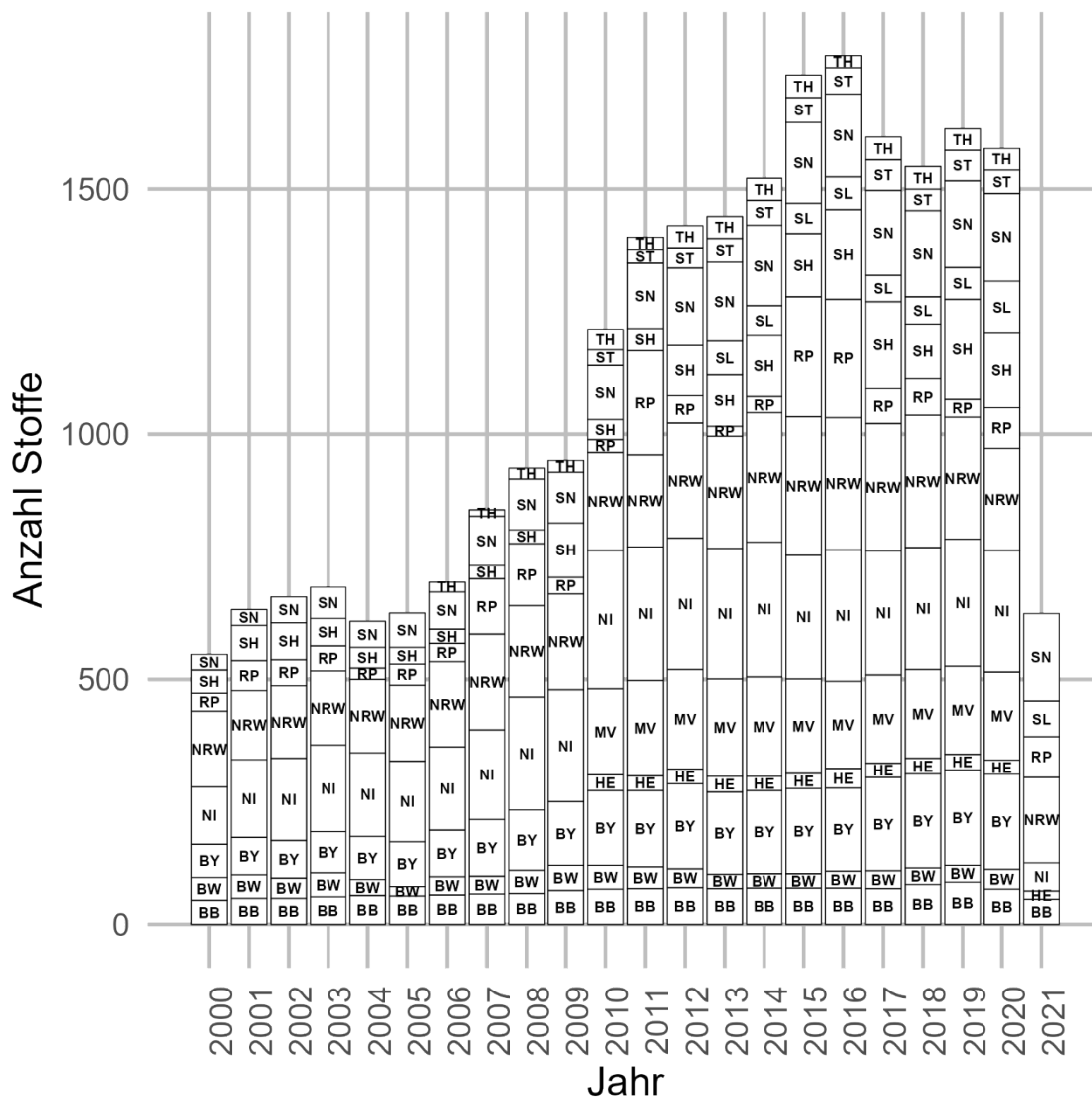
Abbildung 13: Anzahl der gesuchten Stoffe pro Jahr und Bundesland ab 1973 bis 1999.



Quelle: eigene Darstellung BGR.

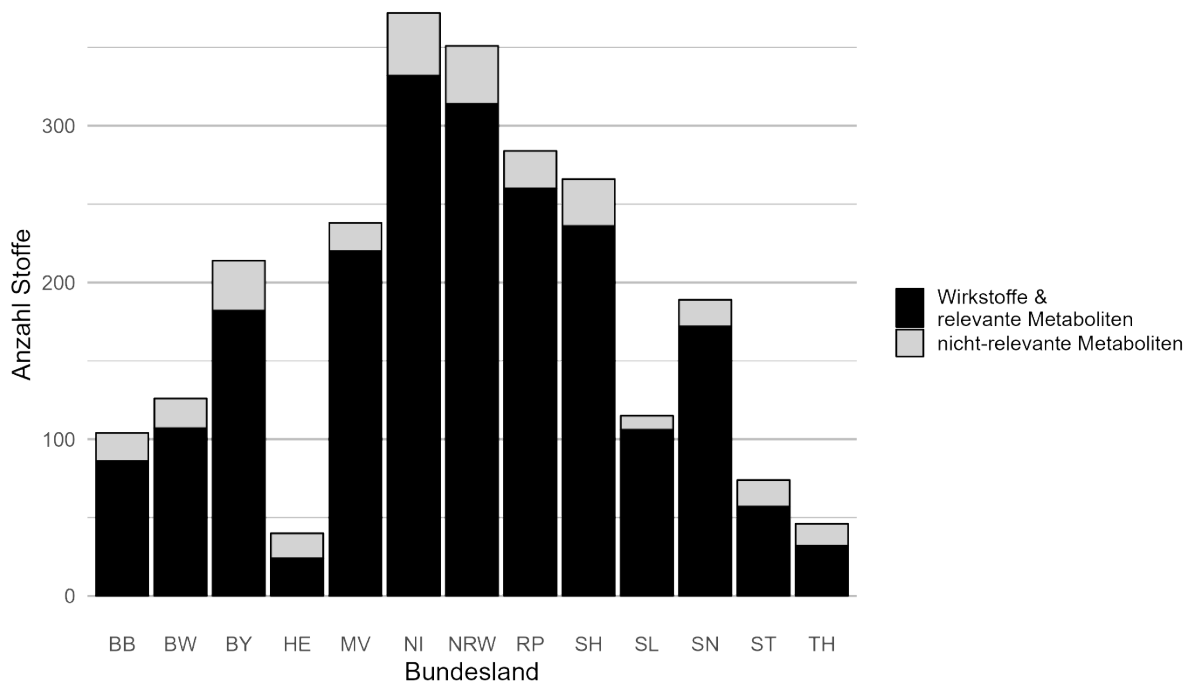
Überschneidungen der Stoffspektren wurden beibehalten, d.h. Stoffe, die in mehreren Bundesländern gemessen werden, führen dazu, dass die Gesamtanzahl an Stoffen höher erscheint.

Abbildung 14: Anzahl der gesuchten Stoffe pro Jahr und pro Bundesland ab 2000. In den Jahren nach 2010 scheint sich das Messprogramm in den Bundesländern weitestgehend etabliert zu haben.



Quelle: eigene Darstellung BGR.

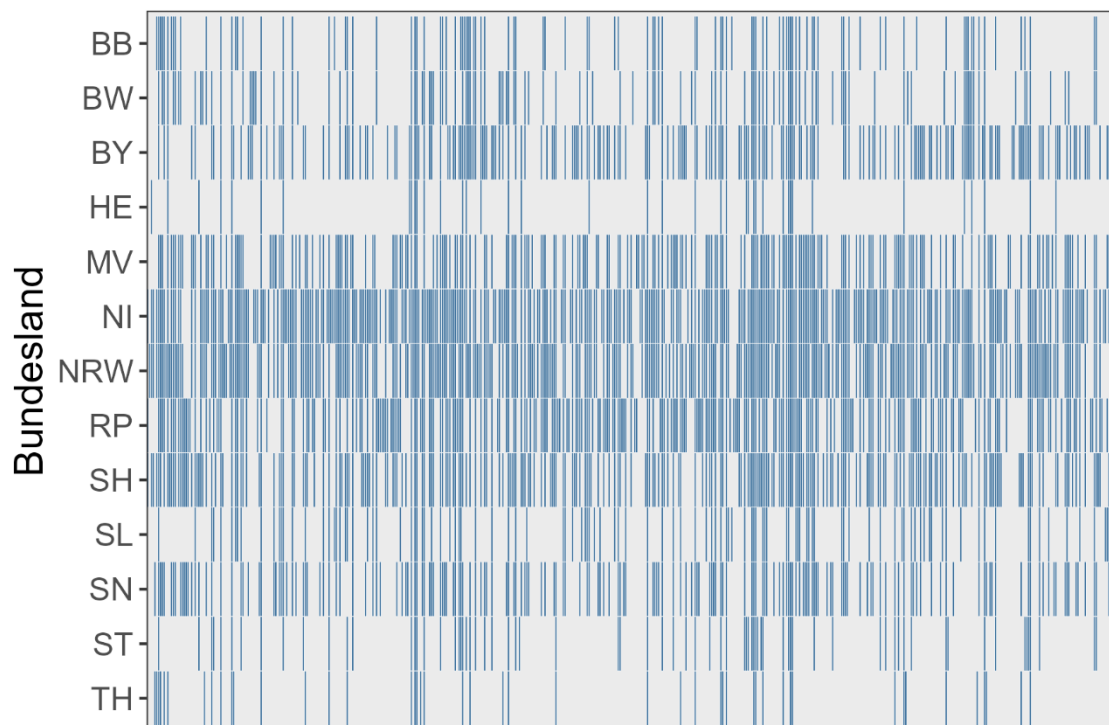
Abbildung 15: Anzahl der im Monitoring beprobten Stoffe aus dem Gesamtdatensatz pro Bundesland.



Quelle: eigene Darstellung BGR.

Nur ein geringer Anteil der Stoffe (meist unter 15%) gehört zu den nicht-relevanten Metaboliten.

Abbildung 16: Barcode-Darstellung des Stoffspektrums pro Bundesland.



Quelle: eigene Darstellung BGR.

Jeder Stoff (Wirkstoffe und Metaboliten) ist als vertikaler Strich dargestellt; die horizontalen Reihen stellen die einzelnen Bundesländer dar.

Im Gesamtdatensatz wurde im Zeitraum 1973-2019 lediglich nach elf Wirkstoffen und sieben Metaboliten in allen 13 Flächenbundesländern beprobt, wovon vier Wirkstoffe zum Stand Juli 2021 als zugelassen galten (sowie vier Metaboliten von noch zugelassenen Wirkstoffen). Es handelt sich ausschließlich um Herbizide bzw. Metaboliten von Herbiziden (Tabelle 3).

Es sind weiterhin insgesamt acht Stoffe, die im Zeitraum 2010-2019 auch in allen 13 Bundesländern mit mindestens einer Detektion über der Bestimmungsgrenze auftraten. Dabei handelt es sich um vier Wirkstoffe und vier Metaboliten, von denen zwei Metaboliten zu aktuell noch zugelassenen Wirkstoffen gehören.

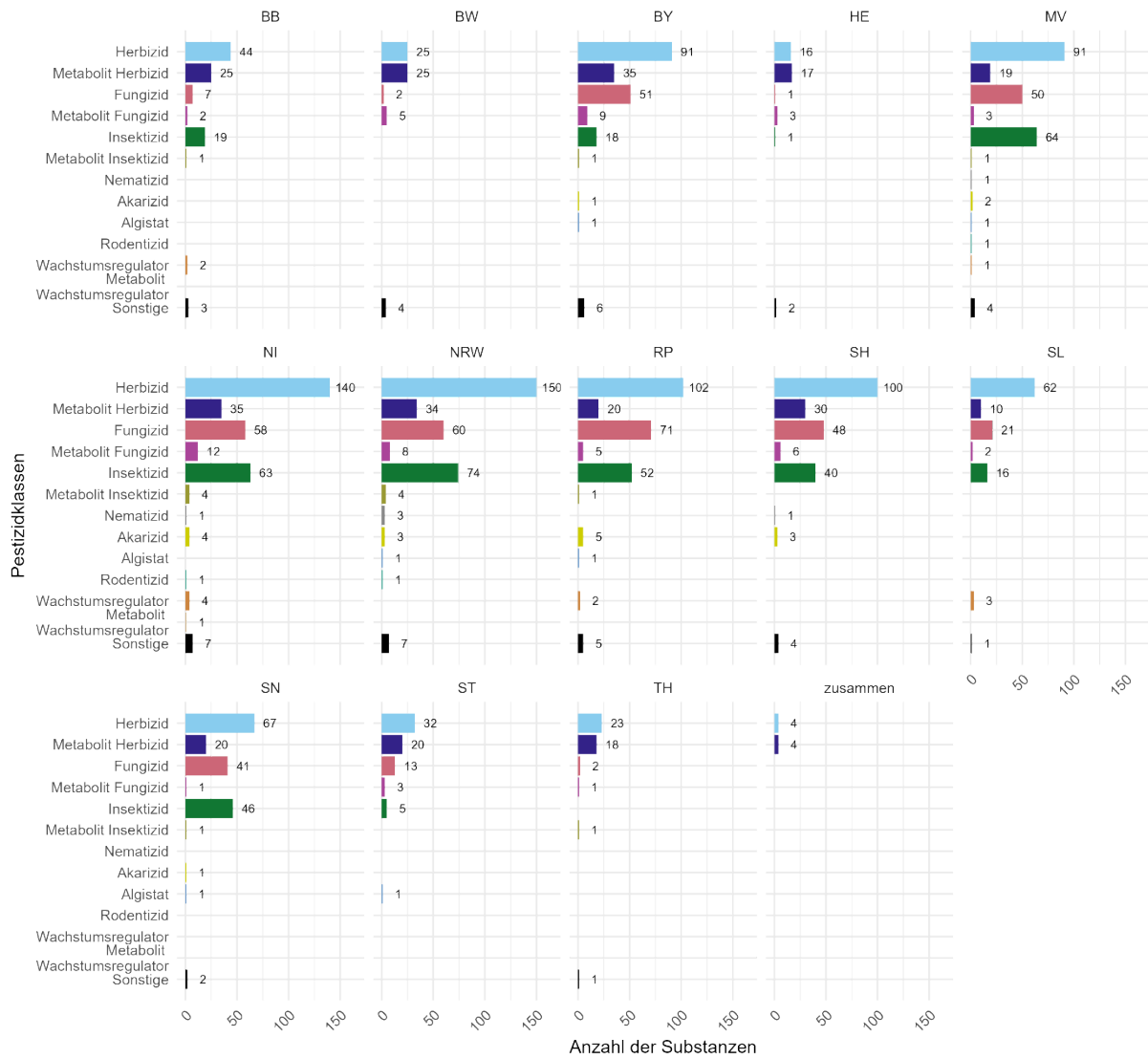
Tabelle 3: Wirkstoffe und Metaboliten, die in allen dreizehn Flächenbundesländern gesucht werden, sind mit ihrer Pestizidklasse und ihrer Zulassungssituation gelistet.

Stoffname	Pestizidklasse	Zulassungssituation (Stand 07.2021)
AMPA	Metabolit Herbizid	Aktuell zugelassen
Atrazin	Herbizid	Nicht mehr zugelassen
Bentazon	Herbizid	Nicht mehr zugelassen
Desethylatrazin	Metabolit Herbizid	Nicht mehr zugelassen
Desethylterbuthylazin	Metabolit Herbizid	Aktuell zugelassen
Desisopropylatrazin	Metabolit Herbizid	Nicht mehr zugelassen
Desphenyl-Chloridazon	Metabolit Herbizid	Nicht mehr zugelassen
Diuron	Herbizid	Nicht mehr zugelassen
Glyphosat	Herbizid	Aktuell zugelassen
Isoproturon	Herbizid	Nicht mehr zugelassen
MCPA	Herbizid	Aktuell zugelassen
Mecoprop	Herbizid	Nicht mehr zugelassen
Metazachlor	Herbizid	Aktuell zugelassen
Metazachlor ESA	Metabolit Herbizid	Aktuell zugelassen
Metazachlorsäure	Metabolit Herbizid	Aktuell zugelassen
Propazin	Herbizid	Nicht mehr zugelassen
Simazin	Herbizid	Nicht mehr zugelassen
Terbuthylazin	Herbizid	Aktuell zugelassen

Zwischen einzelnen oder mehreren Bundesländern kann durchaus eine deutlich größere Überschneidung der Stoffspektren bestehen als bundesweit. Werden beispielweise die Stoffspektren von fünf Bundesländern mit einem relativ breiten Stoffspektrum (NI, RP, MV, NRW, SH) (Abbildung 16) auf Überschneidungen geprüft, dann finden sich 111 Stoffe (96 Wirkstoffe und 15 Metaboliten). Wird diese Auswahl auf Detektionen > LOQ im Zeitraum 2010-2019 gefiltert, dann verbleiben 6 Wirkstoffe (davon 2 zugelassen) und 13 Metaboliten (von 7 zugelassenen Wirkstoffen).

Abbildung 18 verdeutlicht die Anzahl der im Datensatz befindlichen Stoffe je Mobilitäts- bzw. Persistenzklasse. Relevante Metaboliten und potenziell relevante Metaboliten (xM) wurden zusammengefasst. Metaboliten ohne Relevanzstatus wurden als nicht-relevant geführt.

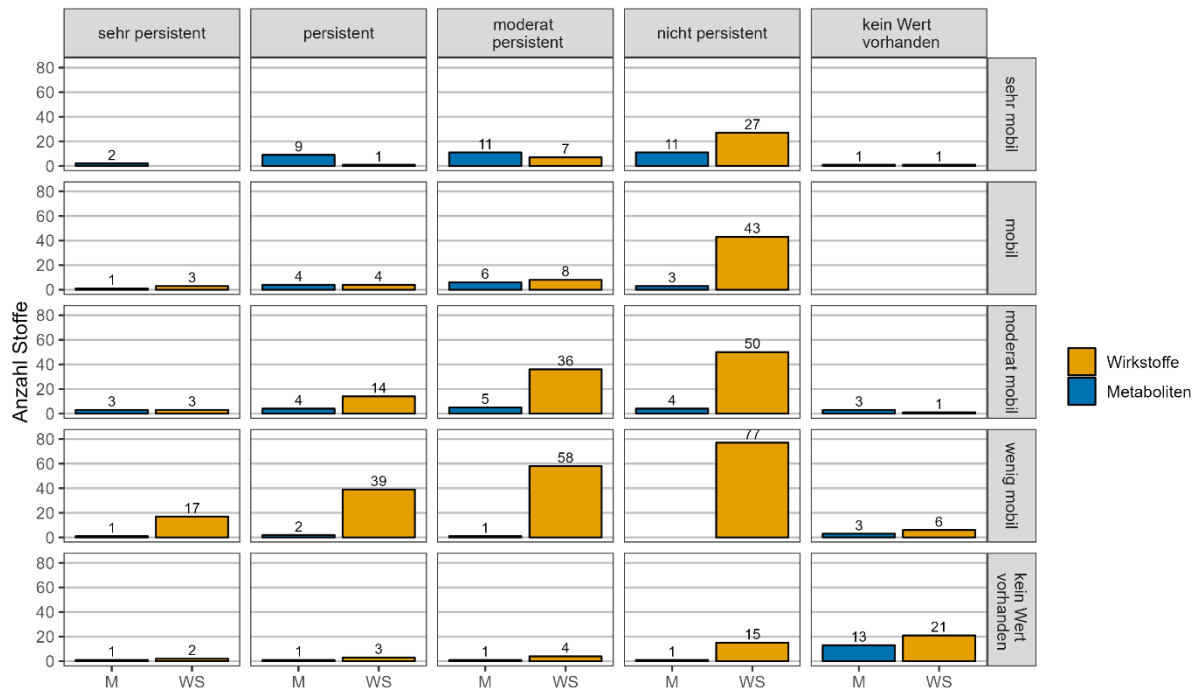
Abbildung 17: Anzahl der Stoffe im Monitoring pro Bundesland und Pestizidklasse, für die Messungen > LOQ vorliegen.



Quelle: eigene Darstellung BGR.

Die größte Gruppe bilden in allen Bundesländern die Herbizide. Zu den Stoffen, die in allen Bundesländern gesucht und mindestens einmal über der LOQ vorgekommen sind, gehören 4 Herbizide und 4 Metaboliten von Herbiziden.

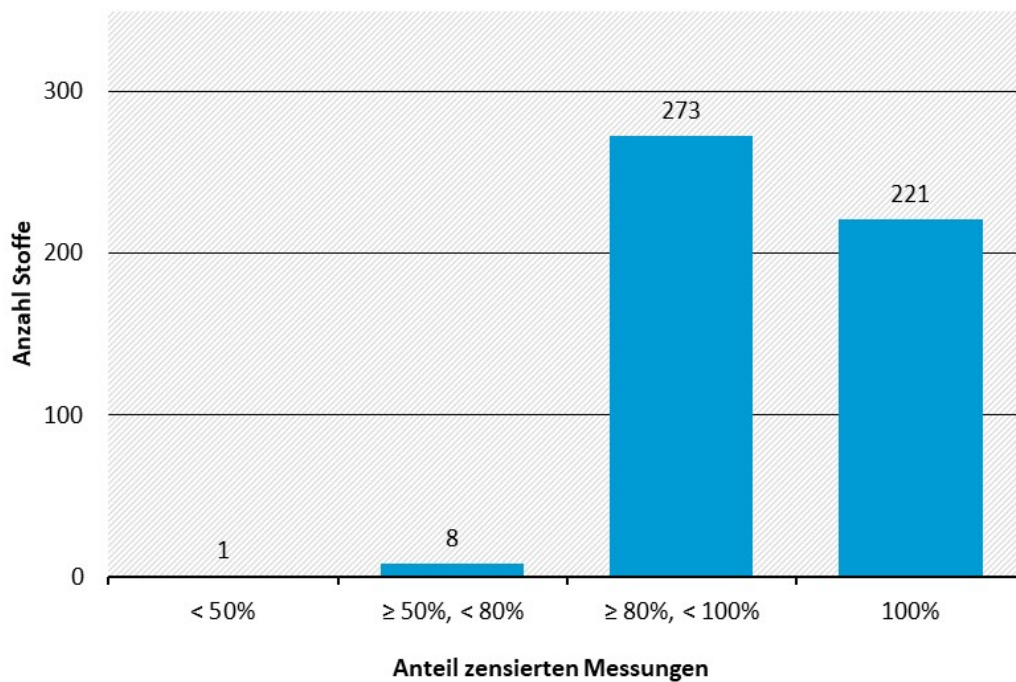
Abbildung 18: Anzahl der Stoffe im Datensatz gruppiert nach Mobilitäts- und Persistenzklasse, gesondert dargestellt für Wirkstoffe und relevante Metaboliten bzw. für nicht-relevante Metaboliten.



Quelle: eigene Darstellung BGR.

Wird der Zensurgrad eines Stoffes nach dem Anteil der Messdaten, die unterhalb der Bestimmungsgrenze fallen, definiert, dann ist dies bei 221 von 530 Stoffen 100% (Abbildung 19). Das heißt, dass bei 43,9% der Stoffe keine Messung oberhalb der Bestimmungsgrenze vorliegt. Mit 278 Stoffen weist ein Großteil einen Zensuranteil zwischen 80 und 100% auf. Neun Stoffe sind zu unter 50% zensiert (Desethyl-Desisopropyl-Atrazin, Desphenyl-Chloridazon, Dimethachlor-Metabolit CGA 369873, Metazachlor ESA, Metazachlorsäure, Methyl-desphenyl-Chloridazon, Metolachlor ESA und der Metabolit NOA 413173). Bei Trifluoressigsäure sind es nur 14,7% zensierte Messungen. Alternativ kann die Anzahl der Messstellen, an der Messungen die Bestimmungsgrenze nicht überschreiten, pro Stoff angegeben werden (Abbildung 20). Hier zeigt sich ein ähnliches Bild, da für einen Großteil der Stoffe an keiner Messstelle, an denen nach diesem Stoff gesucht wurde, die Bestimmungsgrenze überschritten wird. Drei Stoffe haben einen geringeren Anteil an Messstellen, an denen sie nicht über der LOQ detektiert wurden (Trifluoressigsäure, Desphenyl-Chloridazon, Metolachlor ESA). Dementsprechend gibt es 188 Stoffe, für die bei bis zu 25% der Messstellen, an denen nach diesen Stoffen gesucht wird, ausschließlich Funde oberhalb der jeweiligen LOQ vorliegen (Abbildung 21). Bei den folgenden Stoffen sind es 10-20 % der Messstellen: Metolachlor ESA, Metazachlor ESA, Methyl-Desphenylchloridazon, Dimethachlor CGA 369873, Metabolit NOA 413173, Desethyl-Desisopropyl-Atrazin, Terbutylazin SYN 545666, Metolachlor-ca, N,n-dimethylsulfamid und Metazachlorsäure. Diese Analysen müssen jedoch vor dem Hintergrund sehr unterschiedlicher Anzahlen an Messstellen und Messungen pro Stoff verstanden werden.

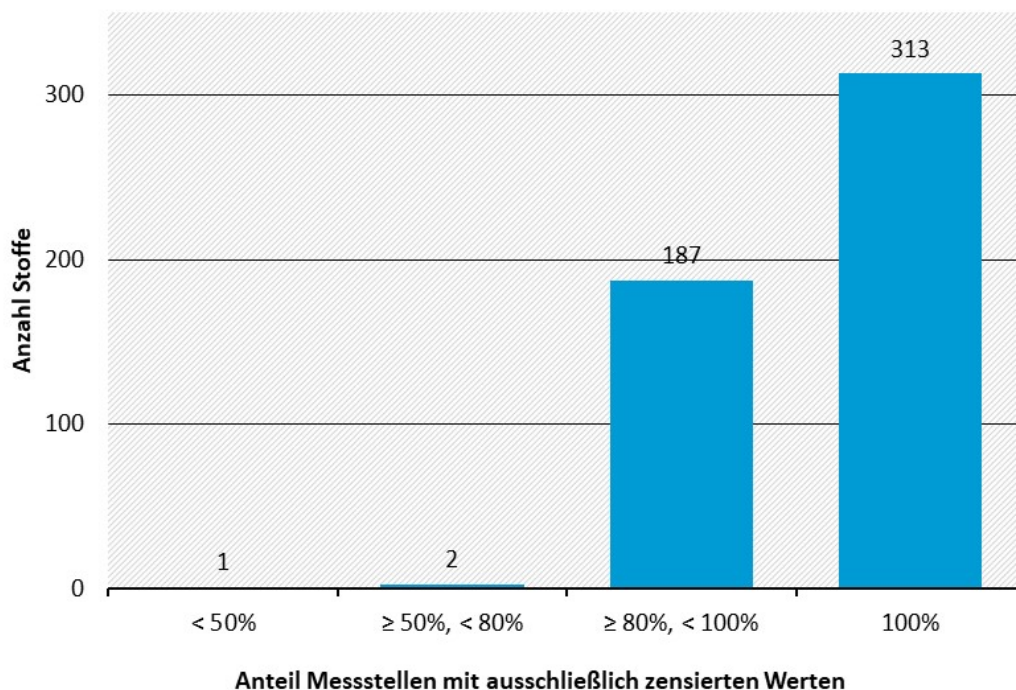
Abbildung 19: Anzahl der Stoffe nach Zensurgrad (Datensatz ab einschließlich 2010).



Quelle: eigene Darstellung BGR.

Nur 9 Stoffe weisen einen Zensuranteil unter 80% auf. 221 Stoffe liegen bei allen Messungen unterhalb der LOQ.

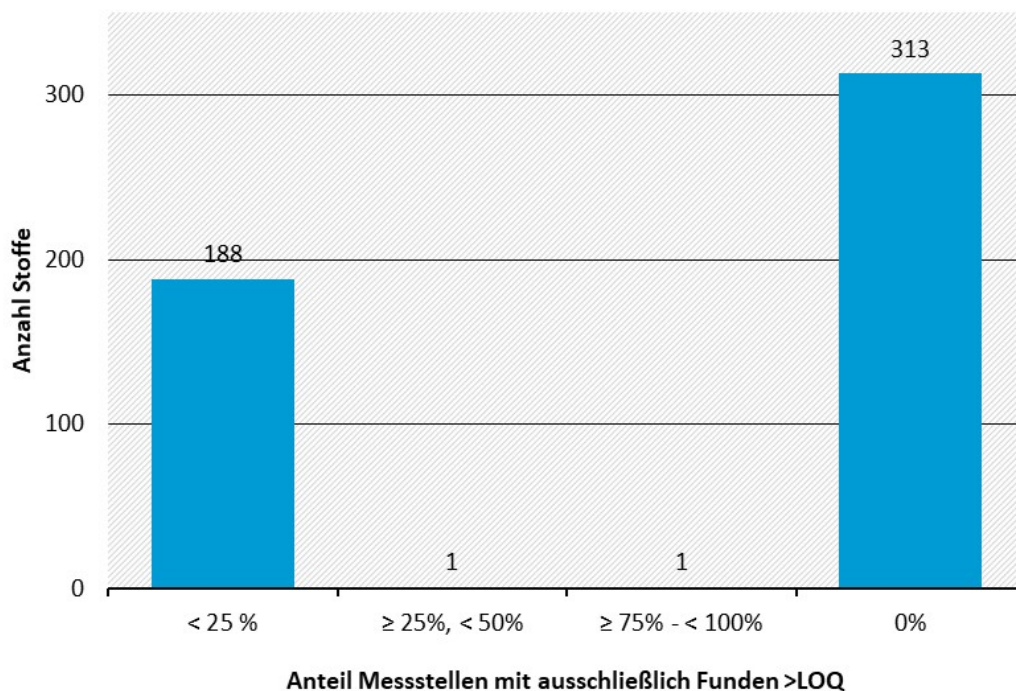
Abbildung 20: Anzahl der Stoffe je Anteil der Messstellen mit ausschließlich zensierten Werten.



Quelle: eigene Darstellung BGR.

313 Stoffe werden an 100% der Messstellen nicht oberhalb der LOQ gefunden. Bei 187 Stoffen sind über 80% der Messstellen ohne quantifizierbare Funde.

Abbildung 21: Anzahl der Stoffe je Anteil der Messstellen, an denen alle Messungen oberhalb der Bestimmungsgrenze liegen.



Quelle: eigene Darstellung BGR.

Bei 188 Stoffen findet sich ein Anteil von bis zu 25% der Messstellen, an denen es ausschließlich zu Funden > LOQ kam.

3.1.3 Fundsituation

3.1.3.1 Gesamtsituation

Betrachtet man die gesamte Datenbank, dann wurden im Gesamtzeitraum Konzentrationen oberhalb der jeweiligen LOQ mindestens einmal für 315 verschiedene PSM-Wirkstoffe und Metaboliten an insgesamt 15.301 Messstellen in allen dreizehn Flächenbundesländern gemessen. Das heißt, dass an 58,4% der Messstellen im Datensatz Wirkstoffe und Metaboliten nachgewiesen werden konnten. Gleichzeitig liegen 96,6% aller im Datensatz enthaltenen Messungen (unabhängig vom Stoff oder der Messstelle) unterhalb der LOQ. Dieser Anteil bleibt auch in späteren Zeitabschnitten relativ konstant bei 97% (bei Betrachtung aller Messungen ab 2010 bzw. ab 2015). Aufgrund der Veränderungen in den Messprogrammen (dem Hinzukommen von Messstellen und Stoffen über die Zeit) sind diese Anteile jedoch nicht direkt miteinander vergleichbar. Im Teildatensatz ab einschließlich 2010 finden sich von 282 Stoffen an 12.228 Messstellen (55,7%) Konzentrationen oberhalb der LOQ, ab 2015 sind es 54,7% der jeweiligen Messstellen.

Betrachtet man die gesamte Datenbank, dann werden an 10.580 Messstellen (40,4%) in allen dreizehn Flächenbundesländern Konzentrationen von $0,1 \mu\text{g L}^{-1}$ mindestens einmal im Gesamtzeitraum überschritten. Insgesamt umfasst dies 188 unterschiedliche Stoffe, bei denen es sich um Wirkstoffe oder Metaboliten handeln kann. Überschreitungen von $1 \mu\text{g L}^{-1}$ ($< 3 \mu\text{g L}^{-1}$) fanden sich für 98 Stoffe an 3.809 Messstellen (14,9%), ebenfalls in allen betrachteten Bundesländern. Insgesamt 66 Stoffe überschritten $3 \mu\text{g L}^{-1}$ ($< 10 \mu\text{g L}^{-1}$) an 1829 Messstellen (7%) in zwölf Bundesländern. Ein Grenzwert von $10 \mu\text{g L}^{-1}$ wurde an 222 Messstellen (0,85%) in zwölf Bundesländern von 39 Stoffen überschritten.

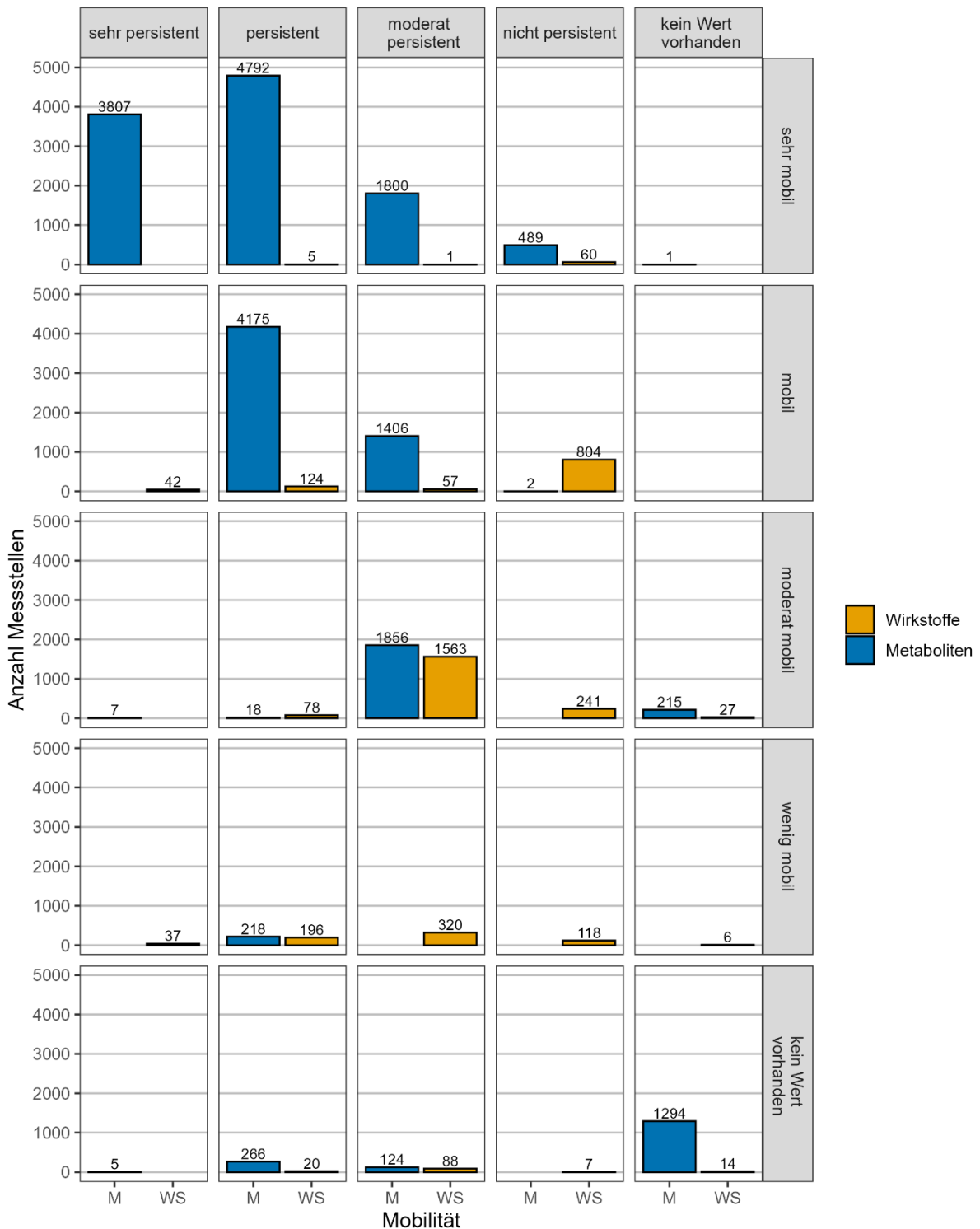
Überschreitungen von $0,1 \mu\text{g L}^{-1}$ von insgesamt 156 Wirkstoffen und relevanten Metaboliten finden sich im Gesamtdatensatz an 6256 Messstellen (19,9%) in dreizehn Bundesländern. Im Zeitraum ab 2010 sind es 129 Wirkstoffe und relevante Metaboliten, die an insgesamt 4.377 Messstellen (23,9% der ab 2010 beprobten Messstellen) den Grenzwert überschreiten.

Der höchste Einzelstoffnachweis findet sich für 4,4-DDT mit $75 \mu\text{g L}^{-1}$ (2005), gefolgt von Metolachlor ESA (2011) und Metribuzin (2018) mit jeweils $70 \mu\text{g L}^{-1}$. Die höchste Summe der quantifizierten Messwerte an einer Messstelle zu einem Probenahmedatum beträgt $141 \mu\text{g L}^{-1}$, welche fünf Stoffen beinhalten (nur Stoffe oberhalb der LOQ). Die maximale Anzahl an Stoffen, die gleichzeitig (am gleichen Datum der Probenahme) an einer Messstelle oberhalb der Bestimmungsgrenze gemessen wurden beträgt 105. Diese Auswertungen sind vor dem Hintergrund der Heterogenität der Messprogramme zu betrachten, da die Anzahl der Stoffe, die im Monitoring berücksichtigt werden, zwischen den Messstellen (sowie zwischen Bundesländern und über die Zeit) stark variiert.

Überschreitungen von $0,1 \mu\text{g L}^{-1}$ werden vorwiegend von sehr mobilen bzw. mobilen Metaboliten hervorgerufen (sofern diese Information vorhanden ist), die moderat persistent bis sehr persistent sind (Abbildung 22). Die Anzahl der Messstellen mit mindestens einer Überschreitung von $0,1 \mu\text{g L}^{-1}$ bei Wirkstoffen und relevanten Metaboliten unterscheidet sich zwischen den Bundesländern (Abbildung 23).

Die räumliche Verteilung der Wirkstoffe und relevanten Metaboliten, die im Gesamtzeitraum mindestens einmal den Grenzwert $0,1 \mu\text{g L}^{-1}$ überschreiten, sind in Abbildung 24 dargestellt, analog nur für Wirkstoffe in Abbildung 25. Grundsätzlich finden sich weniger Messstellen, an denen nicht-relevante Metaboliten $3 \mu\text{g L}^{-1}$ (Abbildung 26) bzw. $10 \mu\text{g L}^{-1}$ überschreiten (Abbildung 27) als bei Wirkstoffen und relevanten Metaboliten.

Abbildung 22: Darstellung der Anzahl von Messstellen mit mindestens einer Überschreitung von 0,1 µg L⁻¹ im Gesamtdatensatz, jeweils für Wirkstoffe und Metaboliten und aufgeteilt Mobilität und Persistenz.



Quelle: eigene Darstellung BGR.

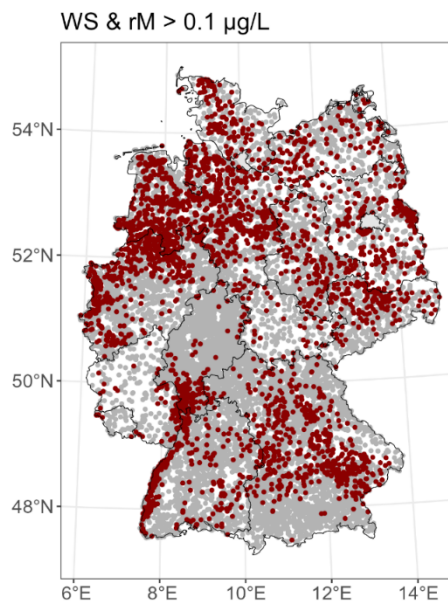
In diese Auswahl fallen nur sehr mobile und mobile, sowie sehr bis moderat persistente Stoffe.

Abbildung 23: Anzahl der Messstellen pro Bundesland mit mindestens einer Überschreitung von $0,1 \mu\text{g L}^{-1}$ bei Wirkstoffen und relevanten Metaboliten im Zeitraum ab 2010.



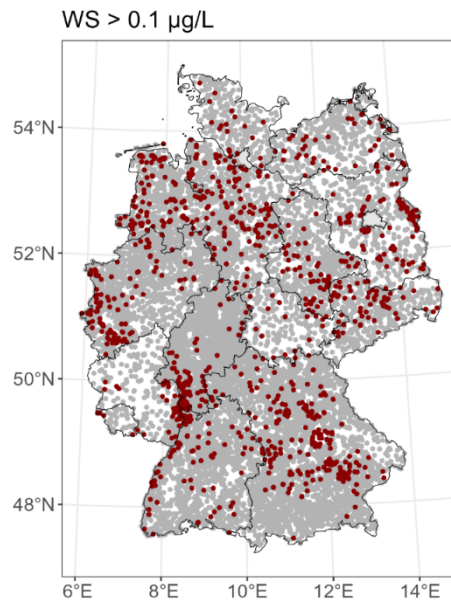
Quelle: eigene Darstellung BGR.

Abbildung 24: Räumliche Verteilung der Messstellen, an denen Wirkstoffe und relevanten Metaboliten $0,1 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.



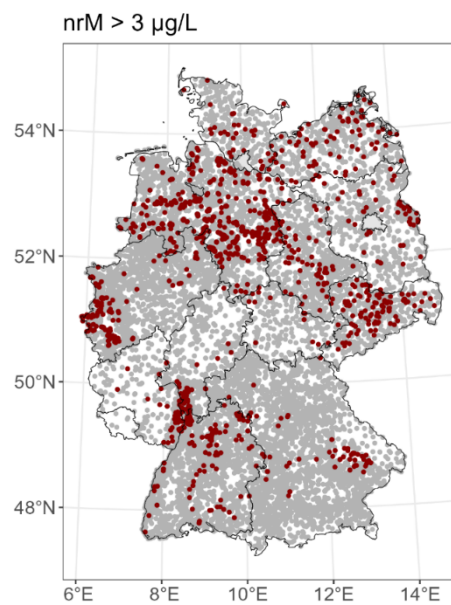
Quelle: eigene Darstellung BGR.

Abbildung 25: Räumliche Verteilung der Messstellen, an denen Wirkstoffe $0,1 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.



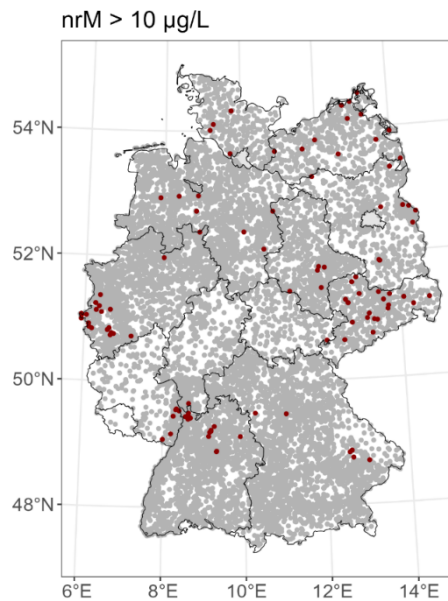
Quelle: eigene Darstellung BGR.

Abbildung 26: Räumliche Verteilung der Messstellen, an denen nicht-relevanten Metaboliten $3 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.



Quelle: eigene Darstellung BGR.

Abbildung 27: Räumliche Verteilung der Messstellen, an denen nicht-relevanten Metaboliten $10 \mu\text{g L}^{-1}$ mindestens einmal überschreiten (rot). Die übrigen Messstellen sind in grau gekennzeichnet.



Quelle: eigene Darstellung BGR.

3.1.3.2 Stoffliche Auswertung

3.1.3.2.1 Stofffranking nach Fundsituation

Eine Einordnung der Stoffe in einer Rangordnung anhand ihrer Fundsituation kann auf verschiedene Weise erfolgen. Ein Ranking anhand der Anzahl der Messwerte oberhalb der LOQ würde aufgrund der Heterogenität der Bestimmungsgrenzen die Rangfolge verzerren. Um Vergleichbarkeit zu gewährleisten, ist ein Ranking anhand der Überschreitungsfrequenz des $0,1 \mu\text{g L}^{-1}$ Grenzwertes sinnvoll. Hier wiederum kann sich an der Anzahl der Messungen als auch an der Anzahl der betroffenen Messstellen als Referenz genommen werden. Der Fokus auf Messstellen eignet sich für Aussagen über die räumliche Verteilung, sagt allerdings wenig darüber aus, ob es sich um einzelne Funde oder eine langfristige Belastungssituation an einer Messstelle handelt. Ein Fokus auf Messungen wiederum führt dazu, dass die Konsistenz der Überschreitung besser abgebildet werden kann, allerdings auch der Einfluss von „Hotspots“ (d.h. Regionen an denen der Stoff stark vermutet und daher Sondermessprogramme zu diesem Stoff durchgeführt worden sind) die relative Bedeutung eines Stoffes im bundesweiten Kontext überhöhen kann. In Cooke et al. (2024) wurden Wirkstoffen und Metaboliten anhand ihrer Fundhäufigkeit, definiert als der Anteil der Messungen, bei denen $0,1 \mu\text{g L}^{-1}$ überschritten wird, gerankt. Zuvor wurden Wirkstoffe und deren Metaboliten, die auch als Biozid in Verwendung sind, herausgefiltert. Das Ranking wurde für zwei Zeiträume (1973-2009 bzw. 2010-2019) gesondert betrachtet. In der folgenden Version des Rankings im Zeitraum 2010-2019 wurden als Biozid gekennzeichnete Pflanzenschutzmittel beibehalten. Das Stofffranking nach Anzahl der Messstellen kann dem aktuellen PSM-Bericht der Länder (LAWA 2024) entnommen werden. Eine Darstellung für die im FARM-Projekt vorliegenden Daten findet sich in Tabelle 4. Eine Tabelle des Stoffrankings basierend auf Messungen finden sich in Cooke et al. (2024; Anhang).

Tabelle 4: Ranking der Wirkstoffe und Metaboliten nach Anzahl der Messstellen, an denen 0,1 µg L⁻¹ überschritten wurden. Stand der Relevanz der Metaboliten ist März 2023, seither wurden zahlreiche xM als rM eingestuft.

Rang	Wirkstoffe und Metaboliten	Wirkstoff	Anzahl Messstellen mit Überschreitung 0,1 µg L ⁻¹	Anzahl Messstellen	Prozentsatz der untersuchten Messstellen	Prozentsatz aller Messstellen	Relevanz (Stand März 2023)	LAWA-Nr.	CAS
1	Trifluoressigsäure	u.a. Flufenacet	3807	4448	85,59	21,78	nrM	4241	76-05-1
2	Desphenyl-Chloridazon	Chloridazon	3013	10964	27,48	17,24	nrM	4014	-
3	Metolachlor ESA	Metolachlor	2289	10225	22,39	13,10	xM	4333	171118-09-5
4	Metazachlor ESA	Metazachlor	2152	10672	20,16	12,31	nrM	4324	-
5	Methyldesphenyl-Chloridazon	Chloridazon	1450	10933	13,26	8,30	nrM	4015	-
6	Metolachlor OXA	Metolachlor	1375	9991	13,76	7,87	xM	4073	152019-73-3
7	NOA 413173	Metolachlor	1051	6479	16,22	6,01	xM	4307	-
8	Metazachlorsäure	Metazachlor	1042	8768	11,88	5,96	nrM	4071	-
9	CGA 369873	Dimethachlor	1012	7547	13,41	5,79	nrM	4264	-
10	n,n-Dimethylsulfamid	Tolyfluanid	830	8244	10,07	4,75	nrM	4000	4710-17-2
11	Dimethachlor ESA	Dimethachlor	288	9547	3,02	1,65	nrM	4490	-
12	Desethylatrazin	Atrazin	273	16480	1,66	1,56	rM	2234	6190-65-4
13	Dimethenamid-P Metabolit M27	Dimethenamid-p	270	3078	8,77	1,54	nrM	4395	-
14	2,6-Dichlorbenzamid	Dichlorbenil	232	9620	2,41	1,33	nrM	2339	2008-58-4
15	Chlorthalonil-SA	Chlorthalonil	226	3462	6,53	1,29	xM	4070	-
16	CGA 357704	Metolachlor	161	2823	5,70	0,92	xM	4304	-

Rang	Wirkstoffe und Metaboliten	Wirkstoff	Anzahl Messstellen mit Überschreitung 0,1 µg L ⁻¹	Anzahl Messstellen	Prozentsatz der untersuchten Messstellen	Prozentsatz aller Messstellen	Relevanz (Stand März 2023)	LAWA-Nr.	CAS
17	Bentazon	-	140	15950	0,88	0,80	WS	2290	25057-89-0
18	Atrazin	-	117	16468	0,71	0,67	WS	2231	1912-24-9
19	Metalaxyl-CA2	Metalaxyl-m	113	2953	3,83	0,65	nrM	4172	-
20	CGA 368208	Metolachlor	91	2380	3,82	0,52	xM	4303	-
21	Ampa	Glyphosat	84	8799	0,95	0,48	nrM	2138	1066-51-9
22	Desethyl-Desisopropyl-Atrazin	Atrazin	66	1574	4,19	0,38	kA	5003	-
23	Terbutylazin Metabolit SYN 545666	Terbutylazin	58	914	6,35	0,33	nrM	4510	-
24	Dimethachlor-CA	Dimethachlor	51	7126	0,72	0,29	nrM	4075	-
25	Flufenacet- ESA	Flufenacet	48	3547	1,35	0,27	nrM	4158	-
26	Ethidimuron	-	43	8329	0,52	0,25	WS	2276	30043-49-3
27	Metazachlor Metabolit BH 479 ⁻¹²	Metazachlor	42	1875	2,24	0,24	nrM	4398	-
28	Bromacil	-	38	13641	0,28	0,22	WS	2289	314-40-9
29	Desisopropylatrazin	Atrazin	28	15983	0,18	0,16	nrM	2262	1007-28-9
30	Diuron	-	28	16454	0,17	0,16	WS	2230	330-54-1

Erklärung zu Spalte 6: Prozentsatz der Messstellen mit Überschreitungen gemessen an der Anzahl der untersuchten Messstellen: Mit untersuchten Messstellen sind alle Messstellen gemeint, an denen nach diesem Stoff gesucht wird.

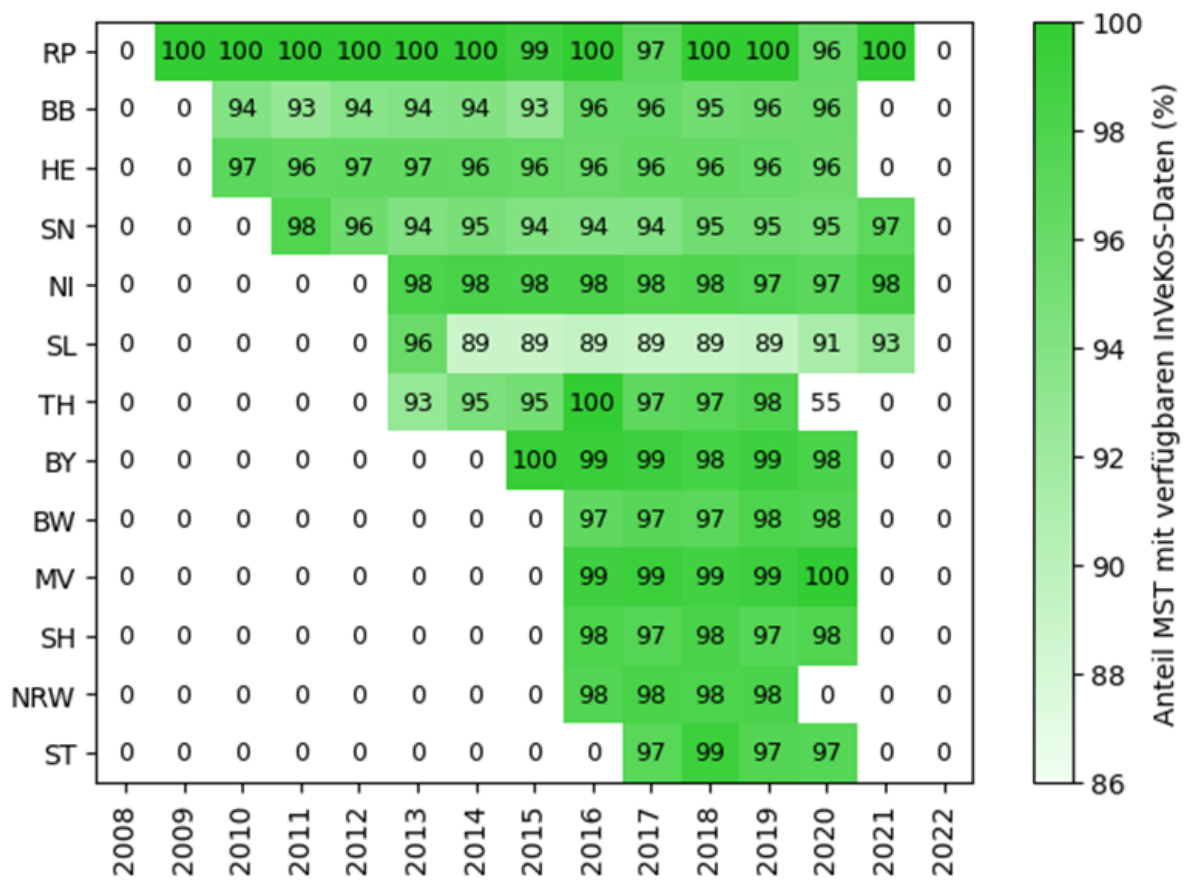
Erklärung zu Spalte 7: Prozentsatz der Messstellen mit Überschreitungen gemessen an der Anzahl aller Messstellen, d.h. alle Messstellen, an denen 2010-2019 nach PSM beprobt wird, unabhängig davon ob es sich um diesen spezifischen Stoff handelt oder nicht

3.1.3.3 Zeitlicher Verlauf

Die zeitliche Auflösung des FARM-Datensatzes ist unter verschiedenen Gesichtspunkten zu betrachten. Dazu zählen (1) die Gesamtverfügbarkeit von Messungen von Wirkstoff- bzw. Metabolitenkonzentrationen über die Jahre unter Berücksichtigung der jeweils verfügbaren Features, (2) die Konsistenz (Regelmäßigkeit und Häufigkeit) der Messungen pro Messstelle (siehe 3.1.1) sowie (3) die Variation der Messwerte an den einzelnen Messstellen über die Jahre.

Die zeitliche Heterogenität und Inkonsistenz der beprobten Messstellen wurden eingehend in 3.1.1 beschrieben. Die zeitlich variable Verfügbarkeit von Features muss ebenfalls berücksichtigt werden, da sie die Benutzbarkeit von Stoffkonzentrationen für maschinelles Lernen einschränken kann. Um ein Modell zu trainieren, müssen alle ausgewählten Features für alle Messungen des jeweiligen Stoffes vorhanden sein. Messungen mit fehlenden Features können nicht berücksichtigt werden. Die InVeKoS-Flächen sind eines der wichtigsten Features und gleichzeitig zeitlich und räumlich begrenzt. In Abbildung 28 ist die Verfügbarkeit von InVeKoS-Daten nach Bundesland und Jahr aufgeschlüsselt. Die frühesten Daten stammen aus Rheinland-Pfalz aus dem Jahr 2009. Brandenburg und Hessen folgen 2010. Die kürzeste Datenverfügbarkeit zeigen Nordrhein-Westfalen und Sachsen-Anhalt mit je vier aufeinanderfolgenden Jahren. Ein bundesweites, zeitlich aufgelöstes Modell ist nur für die Jahre 2017 bis 2019 realisierbar, da in den anderen Jahren für ein oder mehrere Bundesländer keine InVeKoS-Daten zur Verfügung stehen (Abbildung 127). Die Verfügbarkeiten der weiteren Features ist in Abbildung 29 beispielhaft für Dimethachlor-CA aufgeführt. Die Filteroberkante und die grundwasserchemischen Features (insbesondere Nitrat, Eisen und Ammonium) wirken ebenfalls einschränkend. Ihre Verfügbarkeit fluktuiert über Jahre und Stoffe hinweg. Die statischen Features (außer der Filteroberkante) sowie mittlerer Jahresniederschlag und mittlere Jahrestemperatur sind flächendeckend für alle Jahre verfügbar. Da die Features jeweils an unterschiedlichen Messstellen fehlen, ist der Anteil an Messstellen, die alle Features enthalten, sehr gering (siehe Histogramm Abbildung 29).

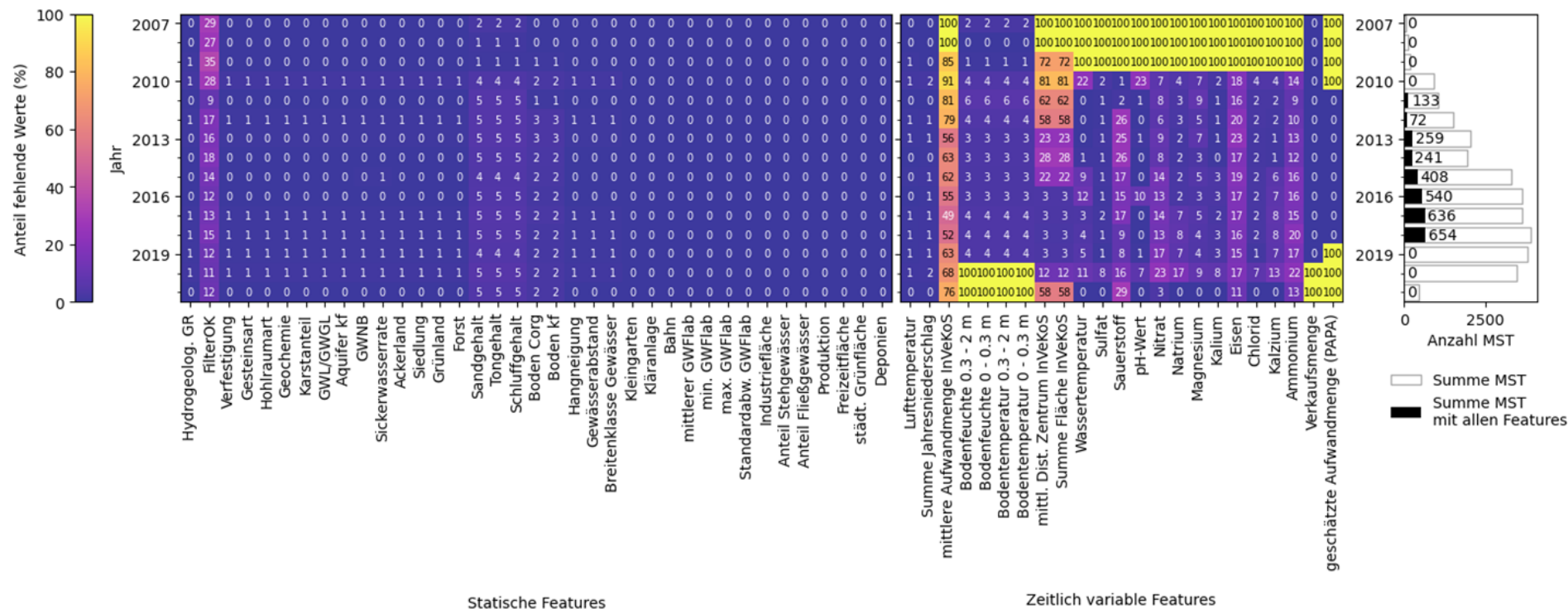
Abbildung 28: Verfügbarkeit von InVeKoS-Daten nach Bundesland und Jahr aufgeschlüsselt. Die Zahlen geben den Anteil an Messstellen mit verfügbaren InVeKoS-Daten in Prozent an.



Quelle: eigene Darstellung BGR.

Die Bundesländer sind nach Start und Länge der Verfügbarkeit sortiert. Rheinland-Pfalz (13 Jahre), Hessen, Sachsen und Brandenburg (je 11 Jahre) haben die beste Datenlage. Die kürzeste Verfügbarkeit zeigen Nordrhein-Westfalen und Sachsen-Anhalt (je 4 Jahre). Nur für die Jahre 2017 bis 2019 stehen für alle Bundesländer InVeKoS-Daten zur Verfügung. Der Anteil der Messstellen mit InVeKoS-Daten liegt bundesland- und jahresübergreifend zu allermeist zwischen 94 und 100%; lediglich im Saarland sind zwischen 2014 und 2019 nur 89% verfügbar und in Thüringen 2020 nur 55%.

Abbildung 29: Verfügbarkeit der Features in Abhängigkeit der Zeit für Dimethachlor-CA. Die Matrizen zeigen den Anteil fehlender Werte je Feature und Jahr. Die linke Matrix enthält die statischen Features; die rechte Matrix enthält die zeitlich variablen Features. Das Histogramm rechts zeigt die Summe aller Messungen der Stoffkonzentrationen pro Jahr als ungefüllte Balken und die Summe der Messungen an Messstellen, für die alle Features verfügbar sind als ausgefüllte Balken.

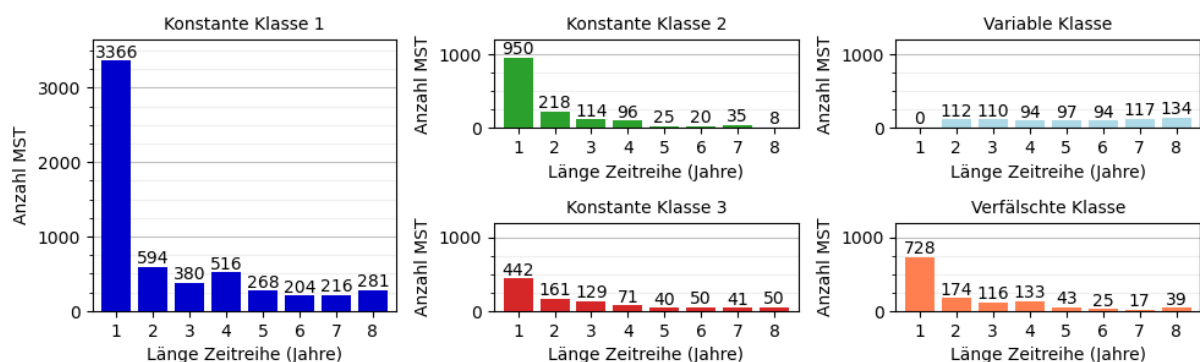


Quelle: eigene Darstellung BGR.

Die Zahlenwerte beziehen sich auf die ausgefüllten Balken. Die statischen Features weisen größtenteils keine oder nur wenige (1,2%) fehlende Werte auf. Nur die Korngrößenanteile Sand, Ton und Schluff haben 4-5% fehlende Werte und die Filteroberkante bis zu 35% fehlende Werte. Die zeitlichen Features sind deutlich lückenhafter. Die grundwasserchemischen Messungen beginnen 2011 und besonders Sauerstoff, Nitrat, Eisen und Ammonium fluktuieren mit bis zu etwa 25% fehlenden Werten. Bodenfeuchte und -temperatur enden 2019. Insbesondere die InVeKoS-Daten zeigen eine hohe Fluktuation durch die unterschiedliche Verfügbarkeit in den einzelnen Bundesländern (siehe Abbildung 26).

Die Variation der gemessenen Konzentrationen an den einzelnen Messstellen ist relevant, um einzuschätzen, wie stark sich zeitliche Trends auf den Lernprozess des Modells auswirken könnten. Aufgrund der Verwendung von Konzentrationsklassen als Zielvariable wirkt sich der genaue Messwert des jeweiligen Stoffes nicht auf das Modell aus, sondern nur die Zugehörigkeit zur jeweiligen Klasse. Dabei ist zu beachten, dass ein Großteil der Messungen ohnehin zensiert ist und keine unterschiedlichen Konzentrationen unterhalb der LOQ aufgelöst werden können. Entscheidend ist also, wie viele Messstellen eine Variabilität der Konzentrationsklasse über die Zeit aufweisen, die vom Modell mit bestimmten zeitlichen Trends in den Features erklärt werden könnte. Abbildung 30 zeigt die Variabilität der Konzentrationsklassen bei zusammenhängenden Zeitreihen beispielhaft für Desphenyl-Chloridazon mit einer umfassenden Feature-Auswahl und einer Einteilung in 3 Klassen (Klassengrenzen 0,05 und 0,5 µg L⁻¹). An der Mehrheit der Messstellen wurden ausschließlich Konzentrationen der Klasse 1 gemessen. Diese können sowohl gänzlich zensiert, teilweise zensiert, als auch gänzlich unzensiert sein. Klassen 2 und 3 enthalten ausschließlich Messungen. Der Anteil an Messstellen, die konstant Klasse 2 oder 3 aufweisen, ist – vor allem mit zunehmender Länge der Messreihe – deutlich geringer. Ebenfalls gering fällt der Anteil an Messstellen aus, bei denen die Konzentration zwischen mindestens zwei Klassen variiert. Einen weiteren Anteil machen Messstellen aus, die verfälschte Konzentrationsklassen aufweisen. Dies bezeichnet Messstellen, bei denen (z.B. durch fehlende Features) Messungen in einzelnen Jahren aussortiert werden mussten, die aber in den aussortierten Jahren eine andere Konzentrationsklasse aufweisen als in den anderen Jahren. Die Bandbreite an möglichen Stoffkonzentrationen an diesen Messstellen ist also – bei der Verwendung der Messungen mit ausreichender Feature-Verfügbarkeit – nicht korrekt abgebildet. Der Anteil dieser verfälschten Messstellen hängt natürlich von der Feature-Auswahl ab und kann mit einem Modell, das auf wenigen Features basiert, potentiell minimiert werden. Im Fall einer Einteilung in 2 Klassen würden einige Messstellen aus der Kategorie ‚variabel‘ verschwinden, da ein Wechsel zwischen Klassen 2 und 3 im Zwei-Klassen-Modell (mit Klassengrenze 0,05 µg L⁻¹) konstant als Klasse 2 eingeordnet wird. Gleichzeitig führt die Zusammenlegung von Klasse 2 und 3 dazu, dass einige Messstellen aus der Kategorie ‚verfälschte Klasse‘ in die Kategorie ‚variable Klasse‘ oder die ‚konstante Klasse 2‘ fallen könnten. Der Anteil an Messstellen mit variabler Konzentrationsklasse ist aber auch diesem Szenario im Verhältnis zur Gesamtzahl der Messstellen sehr klein. Dadurch kann ein zeitliches Modell für die präsentierten Daten nur bedingt zeitliche Trends erlernen.

Abbildung 30: Variation der Konzentrationsklassen bei Desphenyl-Chloridazon in Abhängigkeit der Länge der durchgehenden Zeitreihen.



Quelle: eigene Darstellung BGR.

Die Messstellen sind in fünf Kategorien unterteilt (konstante Klassen 1-3; variable und verfälschte Klasse) und nach Länge der Zeitreihe aufgetragen. Bei den Konzentrationsklassen handelt es sich um 1) $x \leq 0,05 \mu\text{g L}^{-1}$, 2) $0,05 \mu\text{g L}^{-1} < x \leq 0,5 \mu\text{g L}^{-1}$ und 3) $0,5 \mu\text{g L}^{-1} < x$. Die meisten Messstellen sind über alle Jahre der Zeitreihe hinweg in Klasse 1. Einige Messstellen liegen

konstant in Klasse 2 oder 3. Der Anteil an Messstellen mit variablen Klassen ist sehr gering (rund 100 von bis zu 1000). Ein nicht zu vernachlässigender Teil der Messstellen ist ‚verfälscht‘, da durch die Feature-Auswahl oder Unterbrechungen in der Zeitreihe Messstellen mit Konzentrationsklassen aussortiert werden, die in den verbleibenden Jahren nicht repräsentiert sind.

3.1.3.4 Überschreitungsfrequenzen und Verkaufszahlen

Für Metaboliten von Metolachlor, Metazachlor und Chloridazon, die in den Jahren 2010-2019 vollständig oder zeitweise zugelassen waren, wurde der zeitliche Verlauf anhand von jährlichen Überschreitungsfrequenzen von 0,1 und 0,05 $\mu\text{g L}^{-1}$ mit dem Verlauf der bundesweiten Verkaufszahlen verglichen (Cooke et al. 2024). Die unterschiedlichen Messstrategien und damit jährlich wechselnden Messstellen gestalteten sich als der wesentliche die Trendanalyse behindernde Faktor.

3.1.3.5 Der Einfluss von Standortfaktoren auf die Fundsituation

Im Folgenden wird der Einfluss der Standortfaktoren auf die Fundsituation stoffunabhängig als auch stoffspezifisch untersucht. Die Fundsituation wurde hierfür zum einen anhand verschiedener Einteilungen von Konzentrationsklassen und zum anderen als Überschreitungsfrequenz über 0,1 $\mu\text{g L}^{-1}$ beschrieben.

Stoffunabhängig

In Cooke et al. (2024) wurde der Anteil der Messungen in den fünf Konzentrationsklassen (<LOD, < LOQ, < 0,1, > 0,1 und < 1, sowie > 1 $\mu\text{g L}^{-1}$) jeweils für Wirkstoffe und relevante Metaboliten bzw. für nicht-relevante Metaboliten anhand ausgewählter Standortfaktoren und in zwei Zeiträumen untersucht. Hierbei wurde die Verteilung der Konzentrationsklassen hinsichtlich vier Aquifer-Hohlraumarten (Poren-, Kluft-, Kluft-/Poren- und Kluft-/Karst-Grundwasserleiter), drei Landnutzungsklassen (Ackerland/Grünland, Forst und Bebauung) sowie der Tiefe der Filteroberkante unterhalb der Geländeoberkante in fünf Tiefenbereichen (0-2, 2-10, 10-20, 20-50, 50-100 m) dargestellt. Für den Zeitraum 2010-2019 lässt sich zusammenfassend beobachten, dass höhere Anteile von höheren Konzentrationsbereichen in allen Grundwasserleitern vorkommen, allerdings gehäuft in Porengrundwasserleitern (siehe Abbildung 4 in Cooke et al. 2024). Weiterhin verweisen die Ergebnisse auf den Einfluss der Landnutzung in der Umgebung der Messstelle (anteilig mehr höhere Konzentrationsklassen bei Ackerland) und der Filtertiefe (anteilig mehr höhere Konzentrationsklassen bei eher oberflächennahen Filtern).

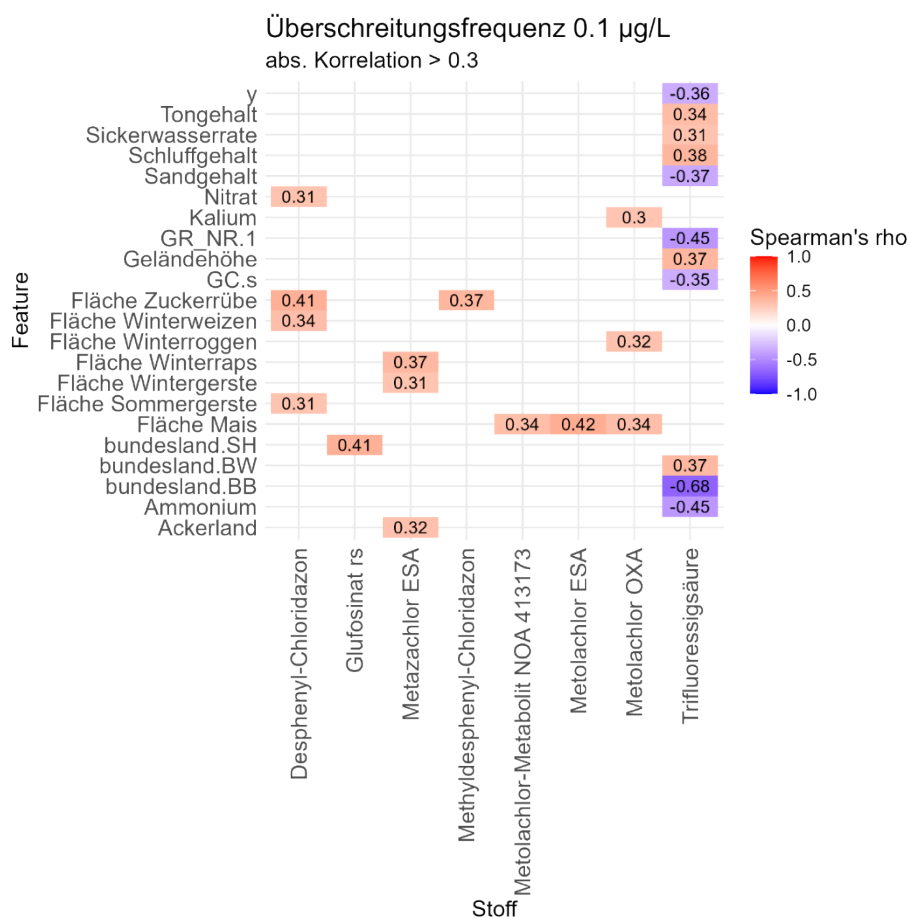
Stoffspezifisch

23,7% der Messstellen weisen mindestens eine Überschreitung des 0,1 $\mu\text{g L}^{-1}$ Grenzwertes für Wirkstoffe und relevante Metaboliten auf. Was haben Messstellen mit Funden von Wirkstoffen und Metaboliten über 0,1 $\mu\text{g L}^{-1}$ bzgl. der Standortfaktoren gemein? In Abbildung 31 werden der Spearman Korrelationskoeffizient ρ für die Korrelationen zwischen allen Standortfaktoren und der Überschreitungsfrequenz von 0,1 $\mu\text{g L}^{-1}$ dargestellt (auch für nrM). Zur verbesserten Lesbarkeit sind nur Korrelationen, die einen absoluten Wert von mindestens 0,3 aufweisen, dargestellt. Sichtbar werden positive Korrelationen zwischen 0,31 und 0,41 zwischen Desphenyl-Chloridazon und der Anbaufläche von Zuckerrübe, Winterweizen und Sommergerste, sowie dem durchschnittlichen Nitratgehalt im Grundwasser. Auch Methylphenyl-Chloridazon zeigt eine leichte, positive Korrelation mit der Anbaufläche an Zuckerrübe auf (0,37). Drei Metolachlor-Metaboliten ESA, OXA und NOA 413173 weisen leichte positive Korrelationen mit der Anbaufläche an Mais auf (0,34 - 0,42). Metolachlor OXA zeigt zudem eine Korrelation von 0,32 mit Winterroggen sowie eine Korrelation mit dem mittleren Kaliumgehalt im Grundwasser (0,30). Metazachlor ESA ist positive korreliert mit Ackerland (0,32) und der Anbaufläche an

Winterraps (0,37) und Wintergerste (0,31). Bei einzelnen Stoffen zeigen sich positive und negative Korrelationen mit einzelnen Bundesländern, was im Bias regional unterschiedlicher Messprogramme als auch in unterschiedlicher Anwendung begründet sein kann. Bei TFA finden sich negative Korrelationen mit dem Bundesland Brandenburg, dem hydrogeologischen Großraum 1 (Nord- und mitteldeutsches Lockergesteinsgebiet, welches auch Brandenburg umfasst), silikatischen Gesteinen, dem mittleren Sandgehalt im Boden und dem Breitengrad. Positive Korrelationen finden sich zwischen TFA und dem Ton- und Schluffgehalt der Böden, der Sickerwasserrate, der Geländehöhe und dem Bundesland Baden-Württemberg.

Insgesamt lässt sich feststellen, dass die Überschreitungen von $0,1 \mu\text{g L}^{-1}$ nur für wenige Stoffe (vorwiegend Metaboliten) Korrelationen $> 0,3$ mit Standortfaktoren zeigen. Auch die Anzahl an Standortfaktoren, die überhaupt Korrelationen $> 0,3$ aufweisen, ist reduziert. Hierbei fallen wie zu erwarten positive Korrelationen mit den zugehörigen und weiteren, ggf. der Fruchtfolge geschuldeten, Kulturen sowie der mittleren Nitrat- und Kaliumkonzentration im Grundwasser auf. Diese Analyse zeigt, dass für einen Großteil der aufbereiteten Standortfaktoren keine Korrelationen gefunden wurden, was bei der Auswahl der Features zu berücksichtigen ist (2.2.3.5). Gleichzeitig ist nicht auszuschließen, dass auch Wechselwirkungen zwischen mehreren Standortfaktoren eine Rolle spielen können, die mit einer bivariate Korrelationsanalyse allein nicht aufgedeckt werden können.

Abbildung 31: Spearman Korrelationskoeffizient rho für die Korrelationen zwischen Standortfaktoren und der Überschreitungsfrequenz von $0,1 \mu\text{g L}^{-1}$.



Quelle: eigene Darstellung BGR.

3.2 Modelle

Im Rahmen dieses Projektes wurden zahlreiche Modelle getestet, die sich bezüglich der (i) betrachteten Stoffe, (ii) der Feature-Auswahl, (iii) der Konzentrationsklassen (binär bzw. 3 Klassen), (iv) der verwendeten Algorithmen (Neuronales Netz, Random-Forest, Logistische Regression), (v) der zeitlich-räumlichen Auflösung und (vi) der Auswahl der Messstellen unterscheiden. Verschiedene Kombinationen dieser Komponenten wurden getestet und bewertet (siehe auch Cooke et al., 2025):

(i) Betrachtete Stoffe:

Ursprünglich wurde eine Betrachtung von PSM-Stoffgruppen mit vergleichbaren Eigenschaften angestrebt. Aufgrund der in 2.2.3.6 beschriebenen Limitationen der Datenlage wurde jedoch letztendlich vorrangig mit Einzelstoffmodellen und dabei insbesondere mit Metaboliten-Modellen gearbeitet. Letztere konnten durch eine ausreichende Datenqualität robuste Modellergebnisse gewährleisten. In den nachfolgenden Ergebnissen werden zuerst die Metaboliten-Modelle der reduzierten und erweiterten Feature-Auswahl behandelt, in einem weiteren Teilkapitel folgen die Ergebnisse der Wirkstoffmodelle (Einzelstoff- und Stoffgruppenmodelle).

(ii) Feature-Auswahl:

Ein weiterer Fokus liegt auf der reduzierten Feature-Auswahl, da diese trotz der geringen Anzahl an Features gute Ergebnisse erzielen konnte und die Interpretation im Gegensatz zur erweiterten Feature-Auswahl mit weniger Unsicherheiten behaftet ist. In den folgenden Teilkapiteln wird daher nur eine relevante Auswahl der Ergebnisse vorgestellt. Weiterhin wird in den nachfolgenden Kapiteln nur auf die Modelle eingegangen, die mit Landnutzungsdaten aus Fernerkundungsprodukten (Blickensdörfer et al. 2022) gerechnet wurden (nicht mit den aus InVeKoS-Daten abgeleiteten Features, da diese bundesweit nur für einen kurzen Zeitraum zur Verfügung standen; siehe 3.1.3.3).

(iii) Konzentrationsklassen:

Modelle mit 3 Konzentrationsklassen zeigten geringe Modellgüten, insbesondere für die dritte Klasse, sowie nahezu identische Effektkurven zwischen Konzentrationsklasse 2 und 3. Dies war bei allen Metaboliten, unabhängig von der Feature-Auswahl und sowohl bei Random-Forest als auch bei den Neuronalen Netzen zu beobachten.

(iv) Verwendete Algorithmen:

Es wird in erster Linie auf die Ergebnisse der Random-Forest-Modelle eingegangen. Gemeinsamkeiten und Unterschiede der Neuronalen Netze werden vergleichend betrachtet. Die Performance der Logistischen Regression wird ebenfalls vergleichend diskutiert.

(v) Zeitlich-räumliche Auflösung:

Bei allen nachfolgend behandelten Ergebnissen der Methoden des Maschinellen Lernens handelt es sich folglich ausschließlich um räumliche Modelle, die mit einer zeitlich aggregierten Datengrundlage hinsichtlich der Zielvariable als auch der Features arbeiten. Zeitlich-räumliche Modelle wurden zwar für einen Teil der Stoffe getestet, die Qualität und Verwendbarkeit der Ergebnisse wurde allerdings aufgrund der in 3.1.3.3 beschriebenen Schwierigkeiten und der Datenlage geschuldeten Limitationen als gering bewertet und in diesem Bericht nicht vorgestellt. Zeitlich-räumlich aufgelöste Modelle wurden daher nicht weiter berücksichtigt, aber sind für zukünftige Auswertungen bei veränderter Datengrundlage wünschenswert.

(vi) Auswahl der Messstellen:

Bei einem Großteil der Modelle wurde mit dem bundesweiten Datensatz, d.h. mit allen Messstellen, an denen nach einem Stoff gesucht wurde, gearbeitet. Die in 3.2.2 und 3.2.3 vorgestellten Ergebnisse beziehen sich auf dieses Szenario. Im Anhang A.6 wird gesondert auf Modelle eingegangen, in denen sich die Daten auf Teilregionen (Pilotgebiete) beziehen.

3.2.1 Reduzierte Feature-Auswahl

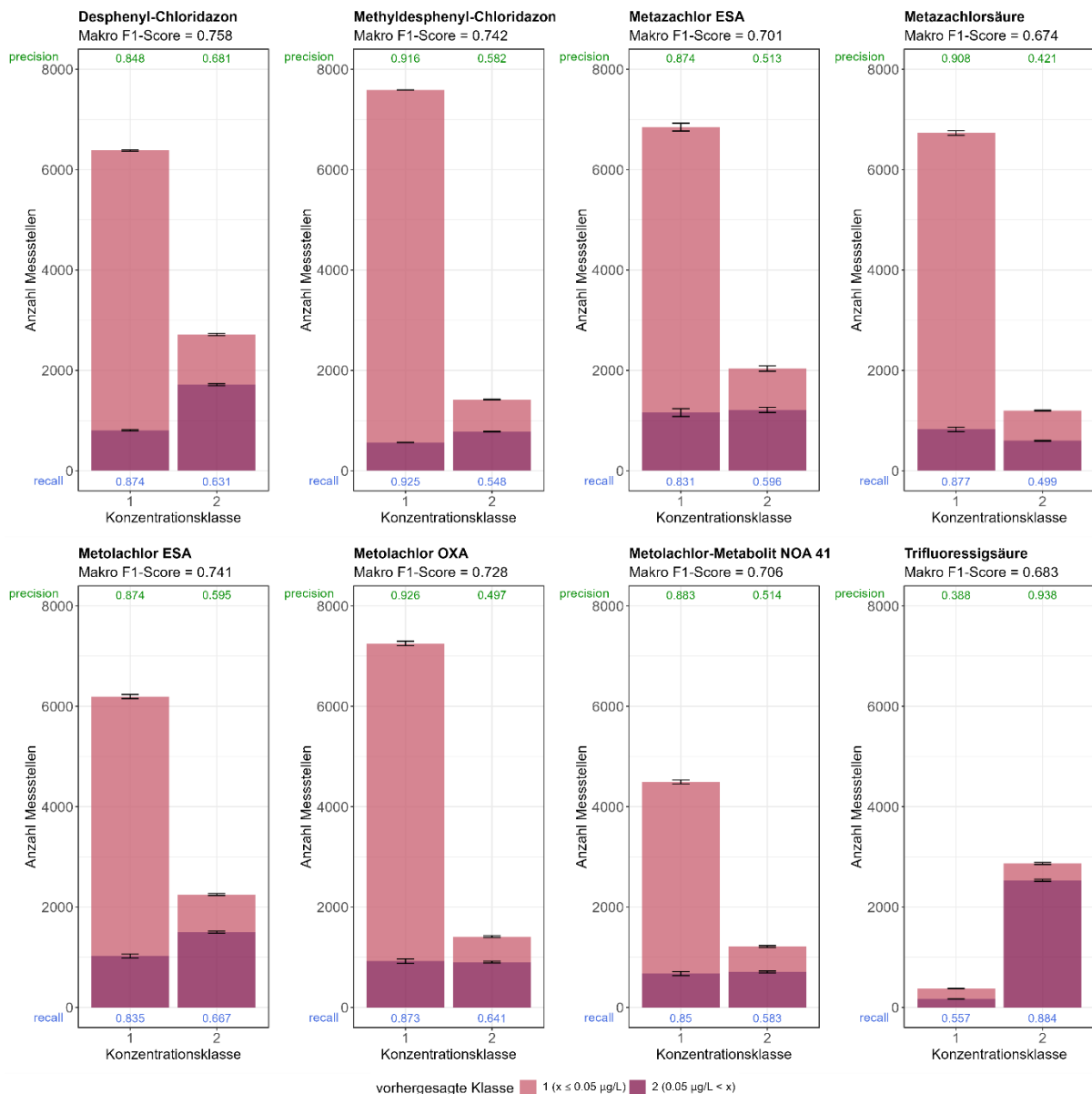
Im Folgenden werden die Ergebnisse der Random-Forest-Modelle der reduzierten Feature-Auswahl der acht Metaboliten vorgestellt (2.2.3.6). Zugunsten der Übersichtlichkeit werden die Ergebnisse der Neuronalen Netze vergleichend angeführt und diskutiert (siehe auch Cooke et al., 2025). Die in nachfolgenden Abbildungen vorkommenden Fehlerbalken stellen bei Random-Forest die Variabilität der unterschiedlichen Startbedingungen (Seeds) und im Falle des Neuronalen Netzes die Variabilität anhand der 5 Teildatensätze (Folds) dar. Die verwendeten Features sind:

- ▶ Flächen pro Kulturklasse, für die der jeweilige Wirkstoff zugelassen ist
- ▶ Tiefe der Filteroberkante unter Geländeoberkante der Messstelle
- ▶ Sandgehalt im Boden
- ▶ Organischer Kohlenstoffanteil im Boden
- ▶ Makroporenklasse vom Boden
- ▶ Langjährige Rate der Grundwasserneubildung
- ▶ Hohlraumart des Aquifers
- ▶ Hydraulische Leitfähigkeitsklasse des Aquifers (siehe Tabelle 8)

3.2.1.1 Vorhersagegüte

Die Macro-F₁-Scores der binären Metabolitenmodelle (Random-Forest; Klassengrenze 0,05 µg L⁻¹) liegen zwischen 0,683 (TFA) und 0,758 (Desphenyl-Chloridazon) (Abbildung 32). In Abbildung 32 werden zudem die unterschiedlichen Verhältnisse der Anzahl der Messstellen in Klasse 1 und 2 bei den acht Metaboliten veranschaulicht. In Bezug auf Recall und Precision wird die erste Klasse durchweg besser vorhergesagt als Konzentrationsklasse 2. Für die meisten Stoffe liegen Precision und Recall der zweiten Konzentrationsklasse über dem Wert einer zufälligen Zuordnung (0,5), bei Metazachlorsäure und Metolachlor OXA liegen diese jedoch leicht darunter. Eine weitere Ausnahme stellt TFA dar. Aus diesem Grund wurde das binäre Modell für TFA mit einer höheren Klassengrenze (0,5 µg L⁻¹) angepasst (Abbildung 33). Mit einem Macro-F1-Score von 0,77 und ähnlich hoher Precision und Recall für beide Konzentrationsklassen wird hier der Einfluss ausgeglichener Klassengrößen auf die Modellgüte deutlich. Die Ergebnisse des Neuronalen Netzes und der Logistischen Regression sind in Tabelle 5 aufgeführt. Der Macro-F1-Score der Random-Forest-Modelle bezieht sich auf den Durchschnitt der verschiedenen Startbedingungen, bei den Neuronalen Netzen wurde der Durchschnitt der fünf Teildatensätze verwendet. Die Macro-F1-Scores der RF- und NN-Modelle sind bei allen Stoffen höher als die der Logistischen Regression, die Unterschiede sind jedoch relativ gering.

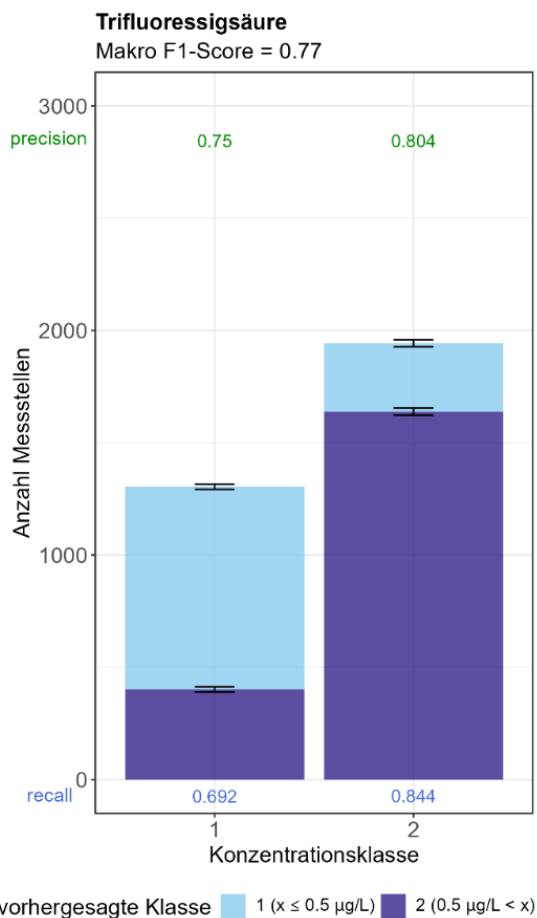
Abbildung 32: Die Modellgüte der binären Random-Forest-Modelle der reduzierten Feature-Auswahl für die acht Metaboliten.



Quelle: eigene Darstellung BGR.

Die Y-Achsen zeigt die Anzahl der Messstellen. Die X-Achse zeigt die tatsächliche Konzentrationsklasse an. Die vorhergesagten Konzentrationsklassen sind farbig markiert (Konzentrationsklasse 1 in hell-, Klasse 2 in dunkelrot). Fehlzuordnungen sind daran zu erkennen, dass sich in beiden Balken die Farbe der jeweils anderen Konzentrationsklasse finden. Daraus lassen sich Gütemaße berechnen: Die Precision (grün) und der Recall (blau) werden jeweils für Konzentrationsklasse 1 und 2 dargestellt. Oberhalb der Plots wird jeweils der Makro-F1-Score angegeben, der sich auf beide Konzentrationsklassen bezieht.

Abbildung 33: In Ergänzung zu Abbildung 32 wird hier die Modellgüte des Random-Forest Modells für Trifluoressigsäure mit der Klassengrenze $0,5 \mu\text{g L}^{-1}$ gezeigt.



Quelle: eigene Darstellung BGR.

Tabelle 5: Modellgüte (Macro-F1-Score) der Logistischen Regression für die binären Metaboliten-Modelle der reduzierten Feature-Auswahl sowie der Neuronalen Netze für die reduzierte und erweiterte Feature-Auswahl.

Algorithmus	Stoffname	Macro-F1-Score	Feature-Auswahl (Anzahl)
Logistische Regression	Desphenyl-Chloridazon	0,69	reduziert
Logistische Regression	Methyldesphenyl-Chloridazon	0,70	reduziert
Logistische Regression	Metazachlor ESA	0,67	reduziert
Logistische Regression	Metazachlorsäure	0,64	reduziert
Logistische Regression	Metolachlor ESA	0,68	reduziert
Logistische Regression	Metolachlor OXA	0,67	reduziert
Logistische Regression	NOA 413173	0,66	reduziert
Logistische Regression	TFA	0,69	reduziert
Neuronales Netz	Desphenyl-Chloridazon	0,72	reduziert

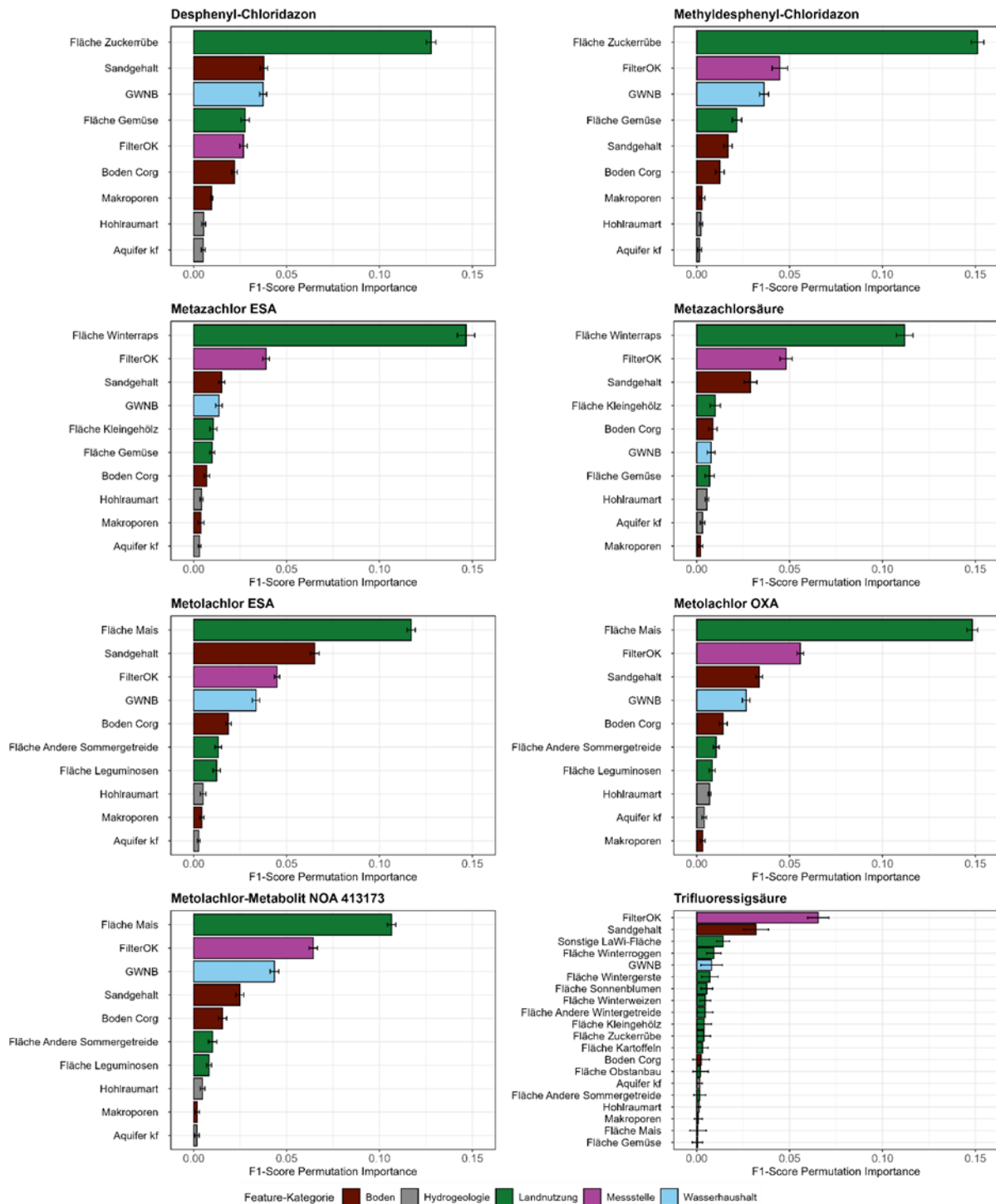
Algorithmus	Stoffname	Macro-F1-Score	Feature-Auswahl (Anzahl)
Neuronales Netz	Methyldesphenyl-Chloridazon	0,73	reduziert
Neuronales Netz	Metazachlor ESA	0,67	reduziert
Neuronales Netz	Metazachlorsäure	0,63	reduziert
Neuronales Netz	Metolachlor ESA	0,71	reduziert
Neuronales Netz	Metolachlor OXA	0,70	reduziert
Neuronales Netz	NOA 413173	0,67	reduziert
Neuronales Netz	TFA	0,71	reduziert
Neuronales Netz	Desphenyl-Chloridazon	0,76	erweitert
Neuronales Netz	Methyldesphenyl-Chloridazon	0,75	erweitert
Neuronales Netz	Metazachlor ESA	0,70	erweitert
Neuronales Netz	Metazachlorsäure	0,65	erweitert
Neuronales Netz	Metolachlor ESA	0,75	erweitert
Neuronales Netz	Metolachlor OXA	0,74	erweitert
Neuronales Netz	NOA 413173	0,73	erweitert
Neuronales Netz	TFA	0,75	erweitert

3.2.1.2 Featuregewichtung (Feature Importance)

Die Feature Importance für die RF-Modelle mit der reduzierten Feature-Auswahl wurde mittels Permutation (Abbildung 34) und Gini impurity (Abbildung 35) ermittelt und die Features nach ihrer Bedeutung absteigend geordnet (2.2.3.7). Methodenunabhängig lässt sich beobachten, dass das mit Abstand wichtigste Feature im Falle von Desphenyl-Chloridazon und Methyldesphenyl-Chloridazon die Anbaufläche an Zuckerrübe im Puffer gewesen ist. Nachrangige Features nach Permutation-Importance sind bei Desphenyl-Chloridazon der Sandgehalt und die Grundwasserneubildung, bei Methyldesphenyl-Chloridazon die Filteroberkante und die Grundwasserneubildung. Nachrangige Features folgen mit etwas Abstand.

Bei Metazachlor ESA und Metazachlorsäure steht die Anbaufläche an Winterraps an erster Stelle, an zweiter Stelle steht die Filteroberkante in beiden Fällen. Bei der Metazachlorsäure folgt an dritter Stelle der Sandgehalt. Weitere, nachrangige Features unterscheiden sich meist nicht signifikant voneinander. Bei den drei Metolachlor-Metaboliten Metolachlor ESA, Metolachlor OXA und NOA 413173 stellt sich die Anbaufläche an Mais im Puffer als wichtigstes Feature heraus. Der Vorsprung zu den nachfolgenden Features ist bei Metolachlor OXA größer. Nachrangige Features sind der Sandgehalt, die Filteroberkante und die Grundwasserneubildung, die relative und absolute Einordnung dieser Features unterscheidet sich zwischen den drei Stoffen. Bei TFA wird der Filteroberkante die größte Bedeutung beigemessen, an zweiter Stelle steht der Sandgehalt. Insgesamt sind die Feature-Importance aller Features deutlich geringer als bei anderen Metaboliten. Nachfolgende Landnutzungsfeatures zeigen relativ geringe Werte und lassen sich in ihrer Rangordnung schwer unterscheiden.

Abbildung 34: Gewichtung der Features nach der Permutation Feature Importance für das binäre Random-Forest Modell für die reduzierte Feature Auswahl.



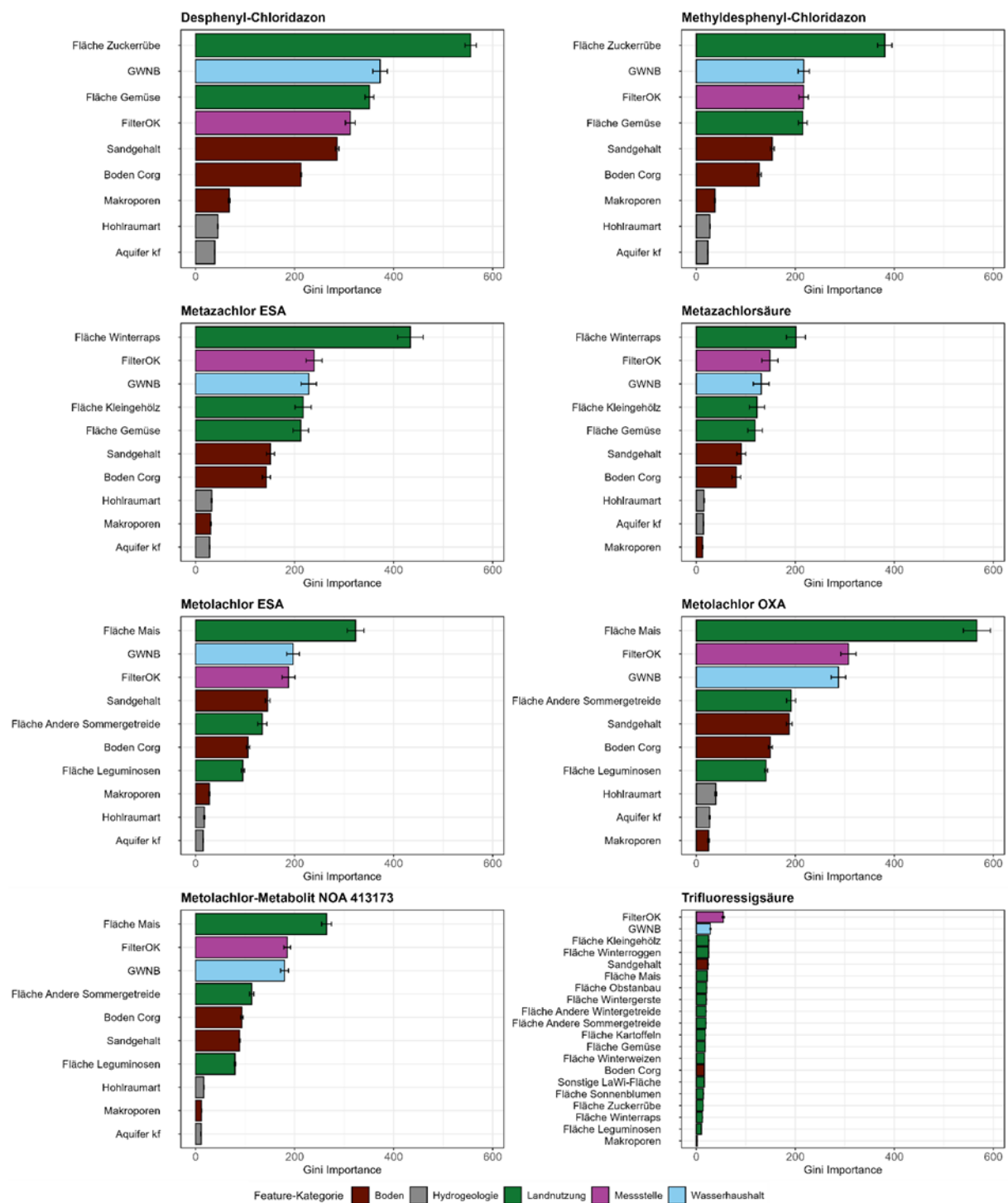
Quelle: eigene Darstellung BGR.

Jede Untergraphik zeigt das Ranking der Features für einen der acht Metaboliten.

Gemäß der Gini impurity sind bei den bereits genannten Metabolitengruppen ebenfalls jeweils die Hauptkulturen an erster Stelle zu finden. Bei den Chloridazon-Metaboliten folgen die Grundwasserneubildung, weitere Kulturen und die Filteroberkante, eine feste Rangordnung lässt sich jedoch schwer ableiten. Bei den Metazachlor-Metaboliten folgen ebenfalls die Filteroberkante, die Grundwasserneubildung und weitere Kulturen, deren Importance sich

jedoch auch überschneidet, sodass eine Bewertung der Reihenfolge bei diesen beiden Features nicht sinnvoll ist. Mit etwas Abstand folgen der Sandgehalt und der organische Kohlenstoff. Bei den Metolachlor-Metaboliten zeigt sich ein grundsätzlich ähnliches Bild. Bei TFA wird nach Gini impurity auch die Filteroberkante als wichtigstes Feature identifiziert, an die zweite Stelle rückt die Grundwasserneubildung, alle weiteren Features weisen ähnliche und geringe Werte auf. Bei beiden Methoden mit Abstand als weniger bedeutend eingestuft wurden die Hohlraumart, die Makroporen und die Durchlässigkeit des Grundwasserleiters.

Abbildung 35: Gewichtung der Features nach der Gini impurity Feature Importance für das binäre Random Forest Modell für die reduzierte Feature Auswahl.



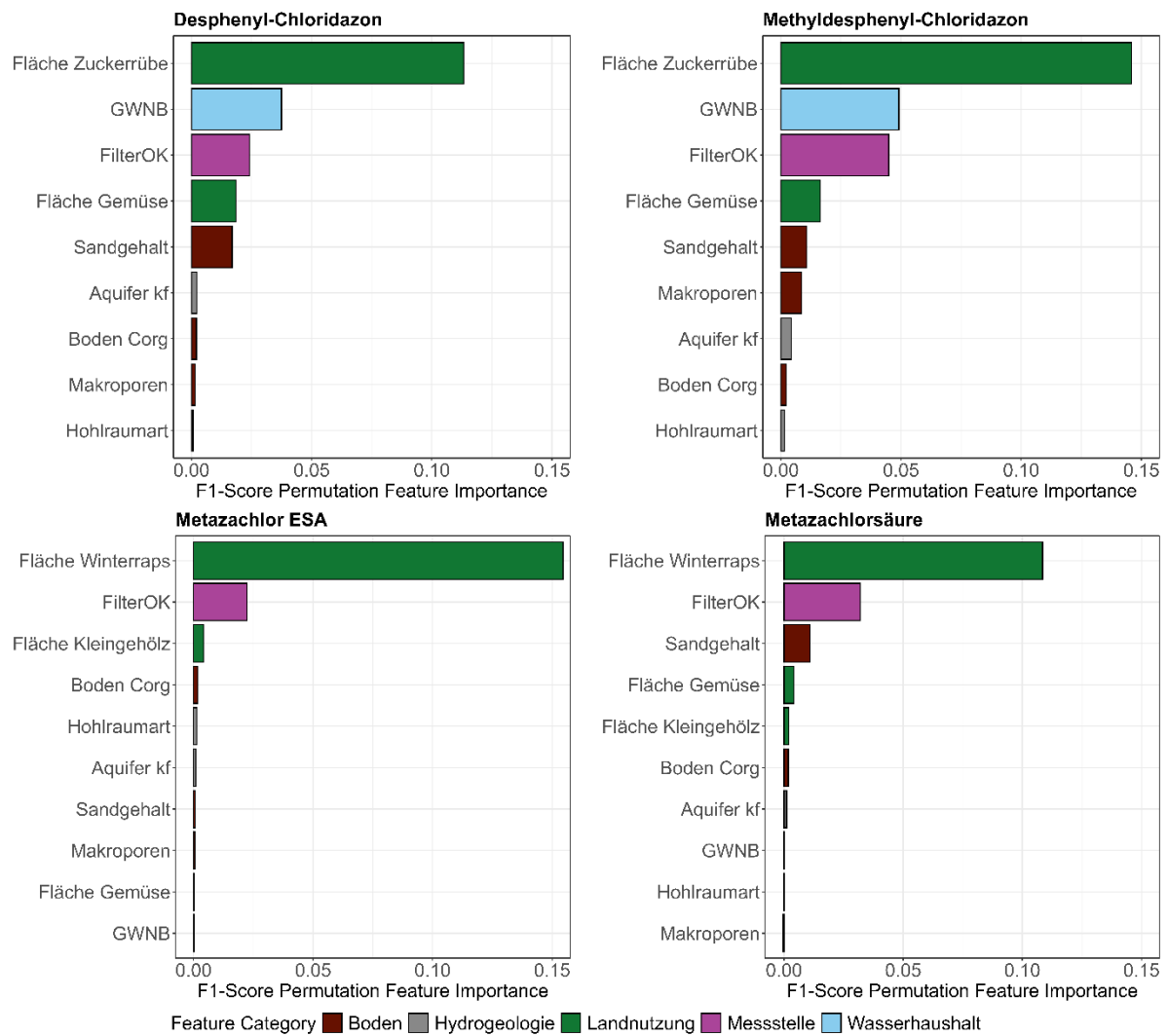
Quelle: eigene Darstellung BGR.

Jede Untergraphik zeigt das Ranking der Features für einen der acht Metaboliten. Diese Graphik erlaubt es, die Feature Importance zwischen den Features und zwischen Stoffen relativ und absolut zu vergleichen.

Die Rangordnung der Permutation-Feature Importance der Neuronalen Netze (Abbildung 36 und Abbildung 37) stimmt in den ersten Features bei allen mit der Rangfolge der Features der Random-Forest-Modelle gemäß Permutation-Feature-Importance oder Gini impurity überein, zum Teil auch bei den zweit- und dritt-wichtigsten Features. Die konkrete Rangfolge

nachfolgender Features mag sich unterscheiden, grundsätzlich stimmt die Einordnung der Feature-Kategorien überein, da Makroporenklasse, die Hohlraumart und der Aquifer-kf meist die letzten Listenplätze einnehmen.

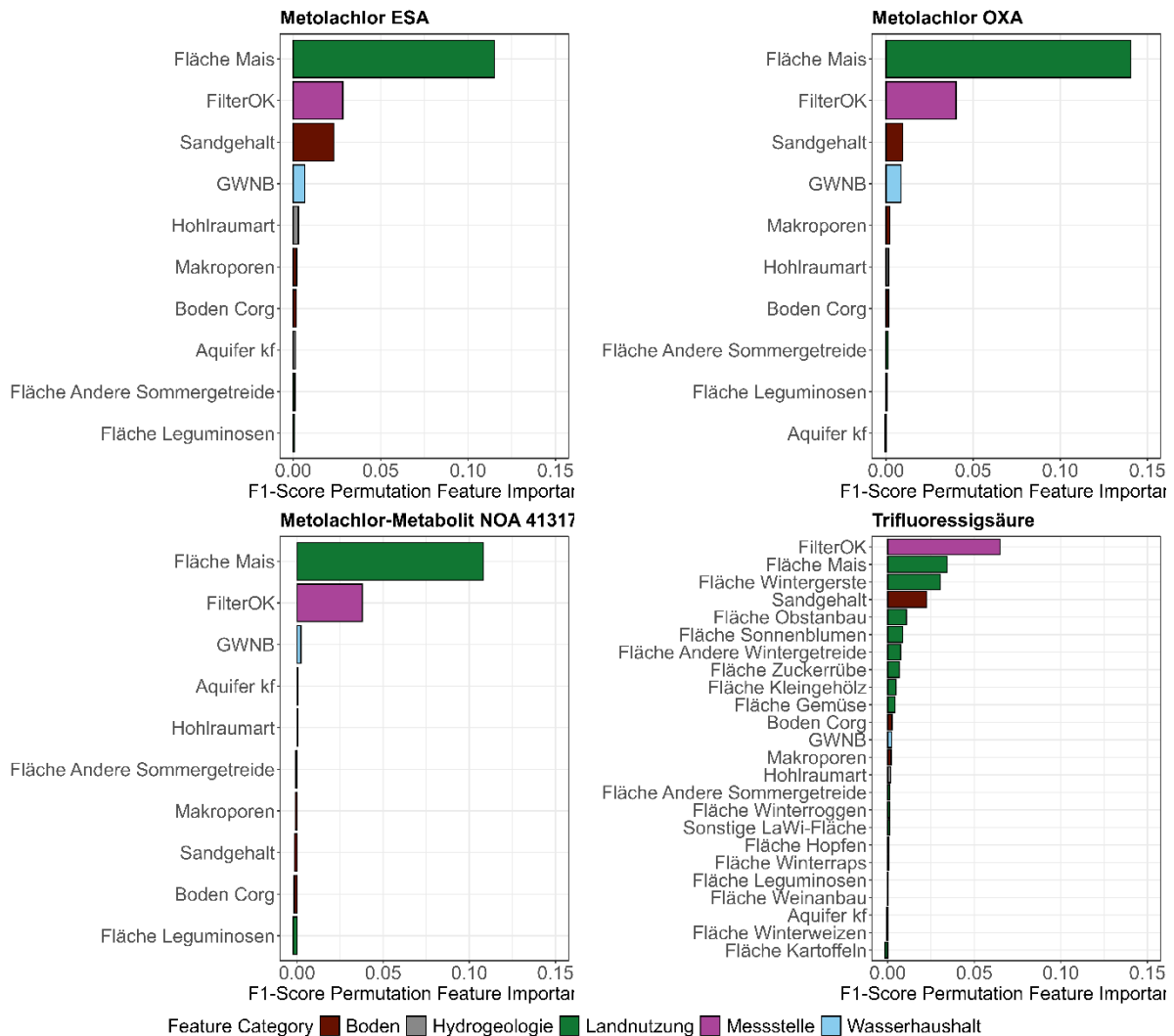
Abbildung 36: Gewichtung der Features nach der Permutation Feature Importance für das binäre Neuronale Netz Modell für die reduzierte Feature Auswahl. Hierbei wird die durchschnittliche Importance, die aus den 5 Teildatensätzen ergibt, gezeigt.



Quelle: eigene Darstellung BGR.

Jede Untergraphik zeigt das Ranking der Features für einen der Metaboliten. Diese Graphik erlaubt es, die Feature Importance zwischen den Features und zwischen Stoffen relativ und absolut zu vergleichen.

Abbildung 37: Gewichtung der Features nach der Permutation Feature Importance für das binäre Neuronale Netz Modell für die reduzierte Feature Auswahl. Hierbei wird die durchschnittliche Importance, die aus den 5 Teildatensätzen ergibt, gezeigt.



Quelle: eigene Darstellung BGR.

Jede Untergraphik zeigt das Ranking der Features für einen der Metaboliten. Diese Graphik erlaubt es, die Feature Importance zwischen den Features und zwischen Stoffen relativ und absolut zu vergleichen.

3.2.1.3 Feature Effekte

Für alle acht Metaboliten und alle verwendeten Features wurden Partial-Dependence Plots erzeugt, die im Folgenden ausgewertet werden. In Abbildung 38 bis Abbildung 45 werden die Feature-Effekte der binären Random-Forest-Modelle (reduzierte Feature-Auswahl) gezeigt. Zugunsten der Lesbarkeit werden die Effekte meist in Bezug auf eine der zwei Konzentrationsklassen beschrieben. Die Effekte der jeweils anderen Konzentrationsklasse zeigen hier in der Regel einen gegensätzlichen Trend, womit die Beschreibung einer Konzentrationsklasse ausreichend ist. Die Y-Achse stellt die Wahrscheinlichkeit der Vorhersage von jeweils Konzentrationsklasse 1 und 2 dar, die X-Achse zeigt den Wertebereich des jeweiligen Features an. Die blaue Linie stellt die gemittelte Effektkurve dar, der hellblaue Schatten verdeutlicht die Variabilität anhand der unterschiedlichen Startbedingungen. Die Effekte der einzelnen Ausprägungen der kategorischen Features sind als Balkenplots dargestellt,

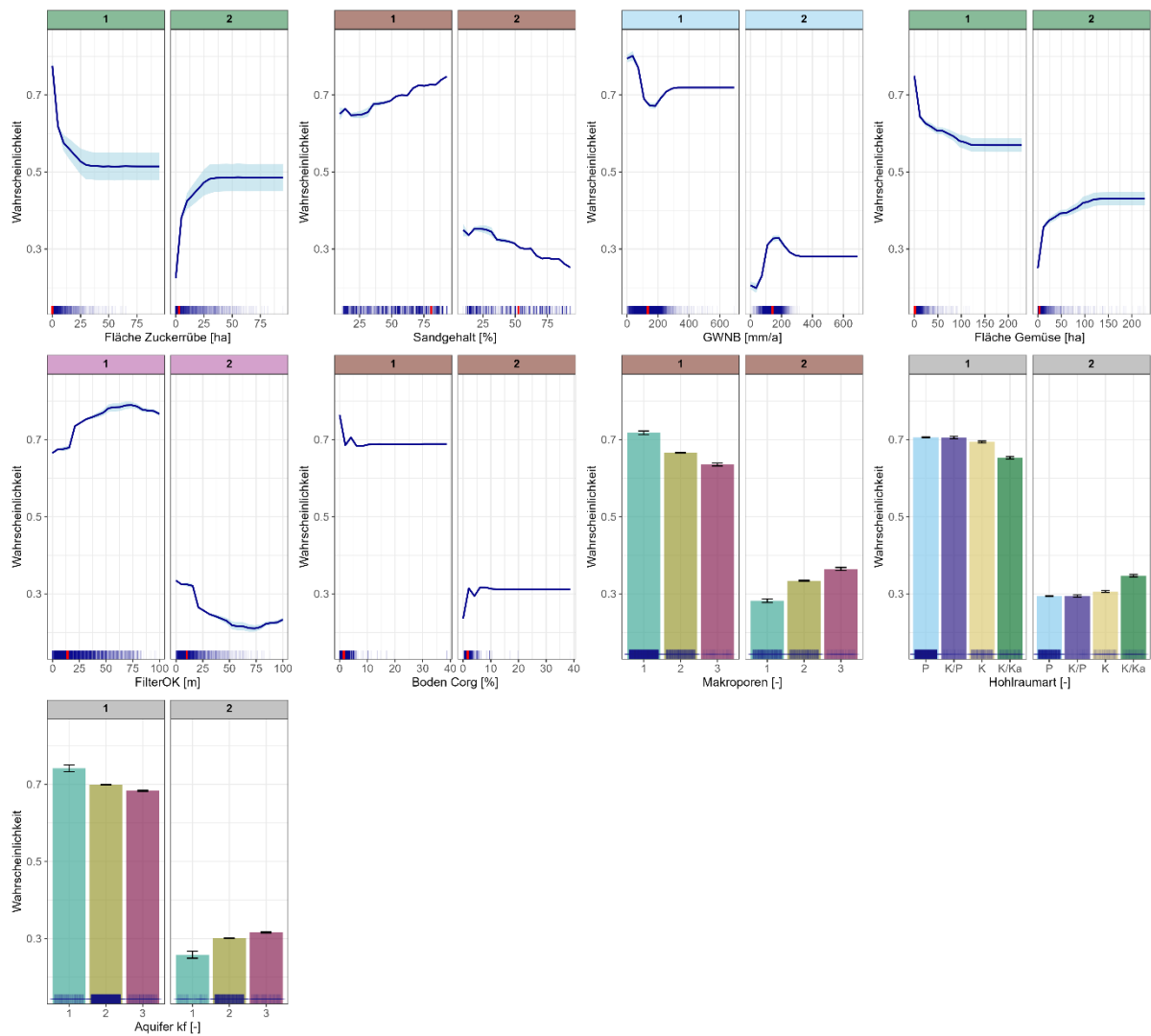
die Fehlerbalken veranschaulichen die Startbedingungen. Die rote Linie verdeutlicht den Median der Feature-Werte.

Chloridazon-Metaboliten

Mit steigender Anbaufläche an Zuckerrübe im Puffer einer Messstelle steigt die Wahrscheinlichkeit der Vorhersage der Konzentrationsklasse 2, während die Wahrscheinlichkeit der Vorhersage der Konzentrationsklasse 1 abnimmt (Desphenyl-Chloridazon, Abbildung 38). Bei steigendem Sandgehalt ist ein steter Abfall der Wahrscheinlichkeit der Konzentrationsklasse 2 zu verzeichnen, Konzentrationsklasse 1 zeigt einen umgekehrten Effektverlauf. Die Grundwasserneubildung führt bei steigenden Werten anfangs zu einer Zunahme der Wahrscheinlichkeit von Konzentrationsklasse 2, ab einem Wendepunkt bei etwa 150 mm/a zeigt sich eine leichte Abnahme der Wahrscheinlichkeit auf die ein Plateau des Effekts folgt. Da nur eine sehr geringe Anzahl an Messstellen höhere Featurewerte aufweist, zeigt sich hier kein Effekt mehr. Die Anbaufläche Gemüse im Puffer zeigt einen vergleichbaren Effektverlauf, wie die Anbaufläche an Zuckerrübe. Bei zunehmender Tiefe der Filteroberkante kommt es zu einer Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 2. Dies ist bis etwa 50 m deutlich, danach flacht die Kurve ab. Beim organischen Kohlenstoff im Boden zeigt sich ein Anstieg der Wahrscheinlichkeit zur Vorhersage von Klasse 2 bei geringen C_{org} Gehalten, der jedoch nicht durchgängig ist, gefolgt von einem Abflachen der Kurve. Bei den Makroporen zeigt sich eine stufenweise Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 mit ansteigender Makroporenklasse (zunehmendem Potenzial der Böden Makroporen auszubilden). Zwischen Poren-, Kluft- und gemischten Grundwasserleitern finden sich keine nennenswerten Unterschiede. Es lässt sich jedoch ein leichter Effekt in Bezug auf eine höhere Wahrscheinlichkeit der Konzentrationsklasse 2 bei Karstaquiferen identifizieren. Bei der hydraulischen Leitfähigkeit (Aquifer-kf-Klasse) zeigt sich eine sehr leichte stufenweisen Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 bei ansteigender kf-Klasse.

Die Feature-Effekte bei Methyl-desphenyl-Chloridazon ähneln grundsätzlich den Effekten, die bei Desphenyl-Chloridazon beschrieben werden konnten. Die Richtung und Stärke der Effektkurven und -balken ist vergleichbar, auch wenn sich die Ausprägungen der Effekte in gewissen Feature-Wertebereichen etwas unterscheiden können (Abbildung 39). Zum Beispiel zeigt sich der abnehmende Trend der Effektkurve der Filteroberkante bei Konzentrationsklasse 2 nur bis etwa 25 m Tiefe. Die Anbaufläche an Zuckerrübe ist auch bei diesen Metaboliten das Feature mit den stärksten Effekten.

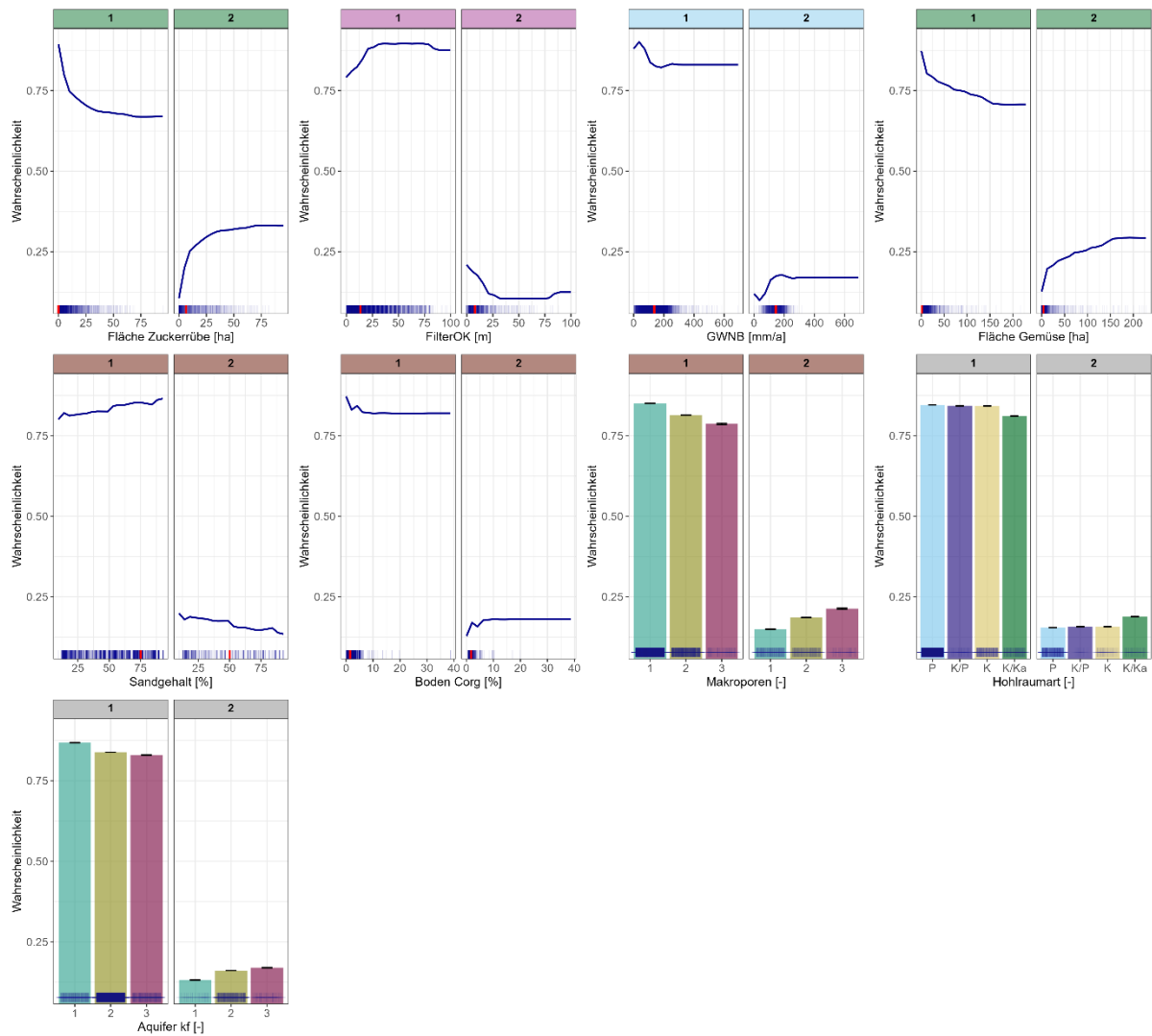
Abbildung 38: Feature-Effekte von Desphenyl-Chloridazon im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen in der Reihenfolge der Feature-Importance erzeugt.



Quelle: eigene Darstellung BGR.

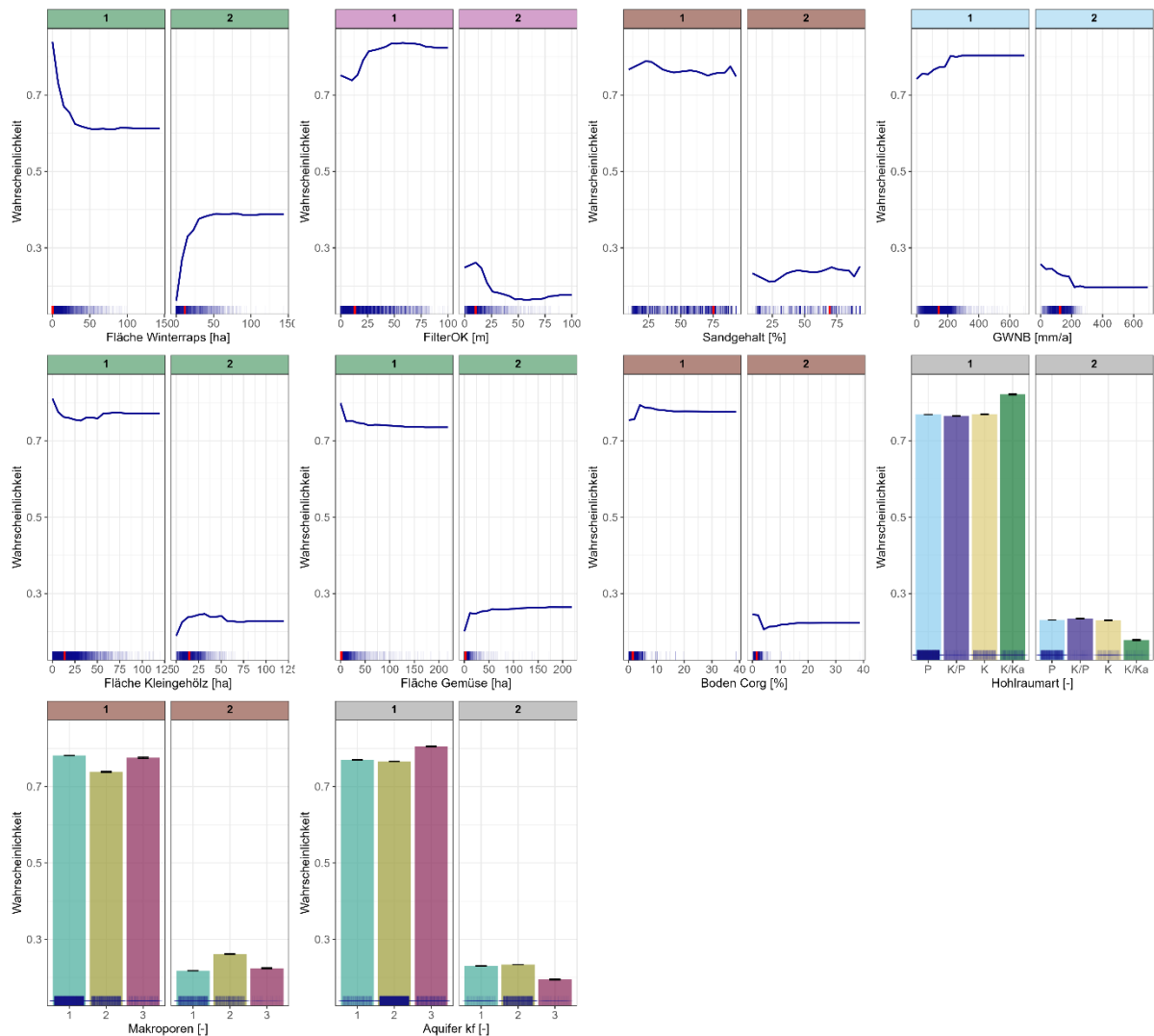
Die Y-Achse stellt die Wahrscheinlichkeit der Vorhersage von jeweils Konzentrationsklasse 1 und 2 dar, die X-Achse zeigt den Wertebereich des jeweiligen Features an. Die blaue Linie stellt die gemittelte Effektkurve dar, der hellblaue Schatten verdeutlicht die Variabilität anhand der unterschiedlichen Seeds. Die Effekte der einzelnen Ausprägungen der kategorischen Features sind als Balkenplots dargestellt, die Fehlerbalken veranschaulichen die Seeds.

Abbildung 39: Feature-Effekte von Methyldesphenyl-Chloridazon im binären Random-Forest Modell der reduzierten Feature-Auswahl.



Quelle: eigene Darstellung BGR.

Abbildung 40: Feature-Effekte von Metazachlor ESA im binären Random-Forest Modell der reduzierten Feature-Auswahl.



Quelle: eigene Darstellung BGR.

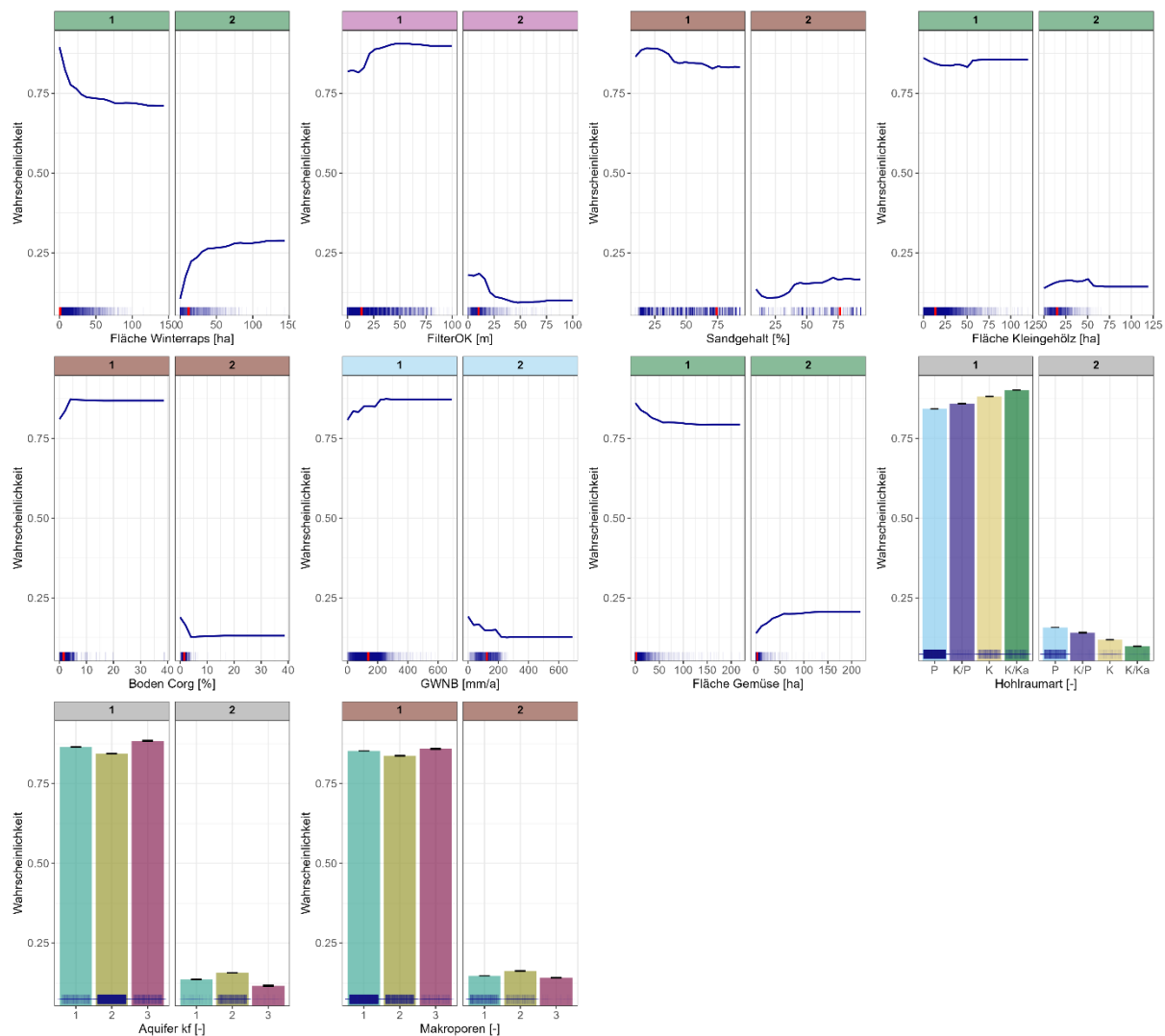
Metazachlor-Metaboliten

Im Fall von Metazachlor ESA ist eine Zunahme der Anbaufläche an Winterraps mit einer Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 1 und einer Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 assoziiert (Abbildung 40). Dies ist auch für die Anbaufläche Kleingehölz als auch für die Anbaufläche an Gemüse sichtbar, jedoch in einem begrenzteren Wertebereich und generell ist der Effekt deutlich flacher. Abgesehen von einer gegenläufigen Tendenz in geringen Tiefen zeigt sich auch bei diesem Metaboliten eine Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 2 bei zunehmender Tiefe der Filteroberkante bis etwa 50 m. Die Effektkurve des Sandgehaltes zeigt Fluktuationen, ist jedoch über den gesamten Wertebereich als flach einzustufen. Somit zeigt das Feature Sandgehalt insgesamt einen zu vernachlässigenden Effekt. Ein Ansteigen der Grundwasserneubildung ist mit einem Abfall der Wahrscheinlichkeit der Konzentrationsklasse 2 verbunden. Eine Abnahme des organischen Kohlenstoffes im Boden ist mit einer Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 2 assoziiert, danach flacht die Kurve ab. Der Effekt der Hohlraumart ist für Poren-, Kluft- und gemischte Grundwasserleiter zu vernachlässigen, bei den Karstaquifern wird eine Zunahme der Wahrscheinlichkeit von Konzentrationsklasse 1 deutlich. Zwischen

Makroporenklasse 1 und 3 sind kaum Unterschiede zu identifizieren, allerdings ist die Makroporenklasse 2 mit einer erhöhten Wahrscheinlichkeit der Konzentrationsklasse 2 verbunden. Eine hohe hydraulische Leitfähigkeit (Aquifer-kf-Klasse 3) verweist auf eine Verringerung der Wahrscheinlichkeit der Konzentrationsklasse 2, Unterschiede zwischen den anderen beiden kf-Klassen sind zu vernachlässigen.

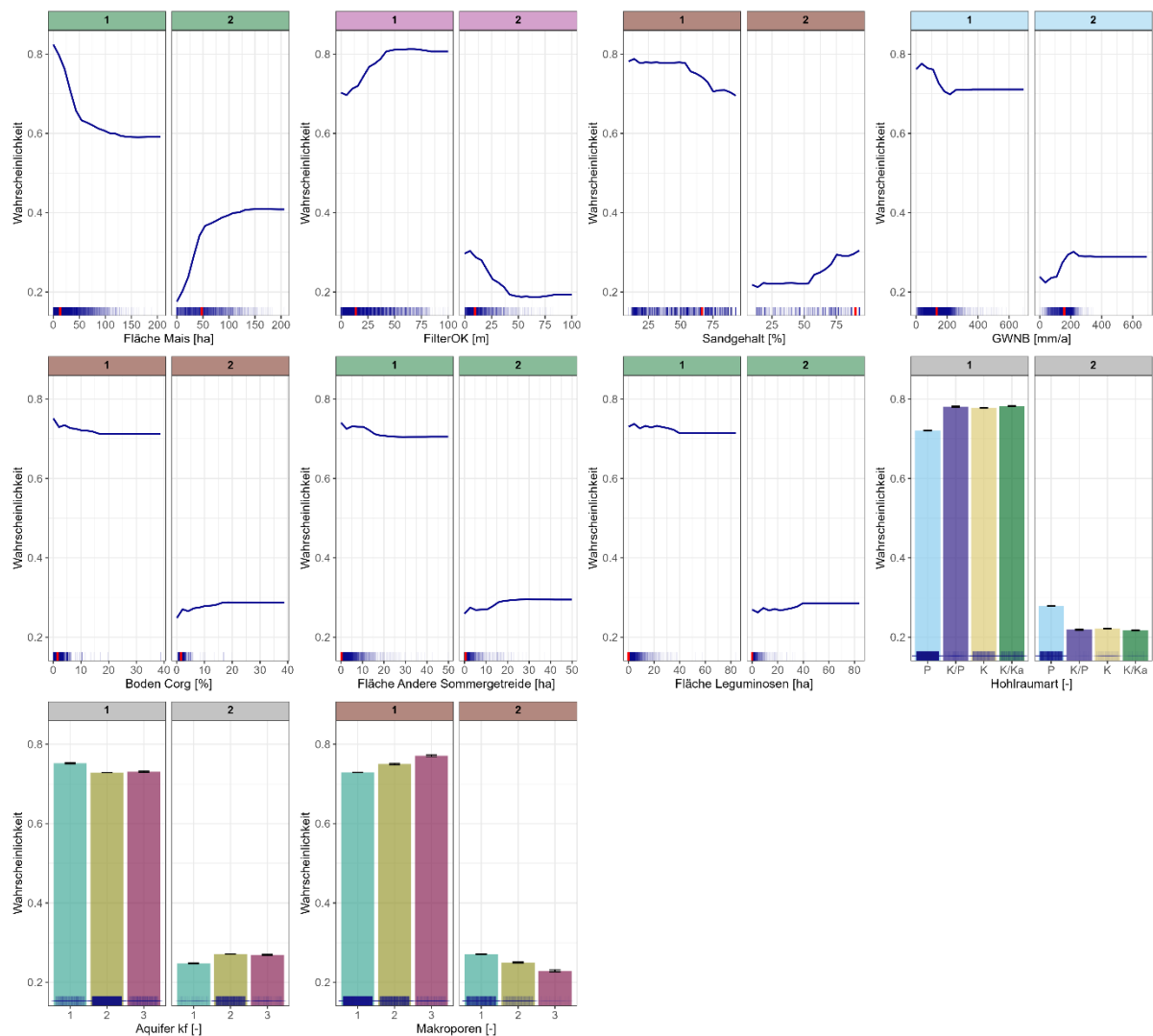
Die Feature-Effekte, die sich bei der Metazachlorsäure (Abbildung 41) zeigen, entsprechen grundsätzlich den bei Metazachlor ESA beschriebenen Effekten. Wiederum sticht die Anbaufläche an Winterraps als bedeutendstes Feature aufgrund starker Effekte heraus. Kleinere Unterschiede sind u.a. beim Sandgehalt sichtbar, der einen sehr leichten Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 aufweist. Der Effekt der Anbaufläche an Kleingehölzen ist zu vernachlässigen. Kleinere Verschiebungen sind auch bei der Hohlraumart und der Aquifer-kf-Klasse zu beobachten.

Abbildung 41: Feature-Effekte der Metazachlorsäure im binären Random-Forest Modell der reduzierten Feature-Auswahl.



Quelle: eigene Darstellung BGR.

Abbildung 42: Feature-Effekte von Metolachlor ESA im binären Random-Forest Modell der reduzierten Feature-Auswahl.



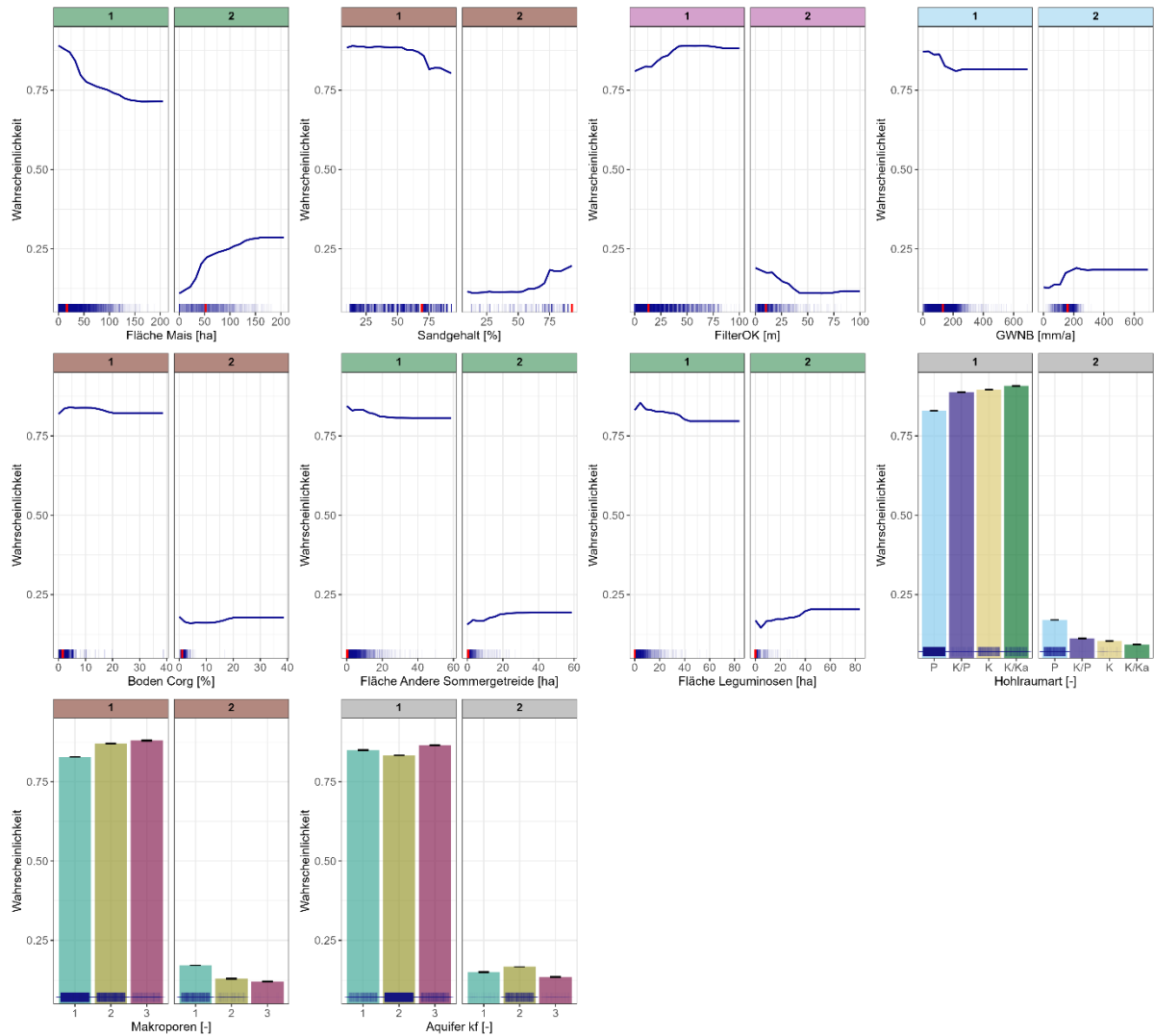
Quelle: eigene Darstellung BGR.

Metolachlor-Metaboliten

Die Anbaufläche an Mais führt bei Metolachlor ESA (Abbildung 42) zu einem Abfall der Wahrscheinlichkeit der Konzentrationsklasse 1 und dementsprechend zu einer Zunahme bei Konzentrationsklasse 2. Der Sandgehalt im Boden zeigt keinen Effekt bis etwa 50%, danach kommt es zu einem positiven Trend mit Bezug auf Konzentrationsklasse 2. Die Filteroberkante verhält sich wie bei anderen Metaboliten und führt mit zunehmender Tiefe zu einer Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 2. Eine steigende Grundwasserneubildung bis 200 mm/a zeigt auch einen leichten positiven Zusammenhang mit der Konzentrationsklasse 2. Features wie der C_{org} Gehalt des Bodens, die Anbaufläche an Sommergetreiden und Leguminosen zeigen zu vernachlässigende Effekte. Ein im Vergleich zu anderen Hohlraumarten relativ erhöhter Anteil an Porengrundwasserleitern ist mit der Konzentrationsklasse 2 assoziiert, während die weiteren Grundwasserleiterarten keine Unterschiede aufweisen. Die Effekte der hydraulischen Leitfähigkeit sind ebenfalls sehr gering, es ist eine leichte Tendenz zu einer höheren Wahrscheinlichkeit der Konzentrationsklasse 1 im Falle der Aquifer-kf-Klasse 1 zu diskutieren. Mit zunehmender Makroporenklasse ist eine Abnahme der Wahrscheinlichkeit der Konzentrationsklasse 2 zu verzeichnen, wobei die Effekte insgesamt nicht sehr stark ausfallen.

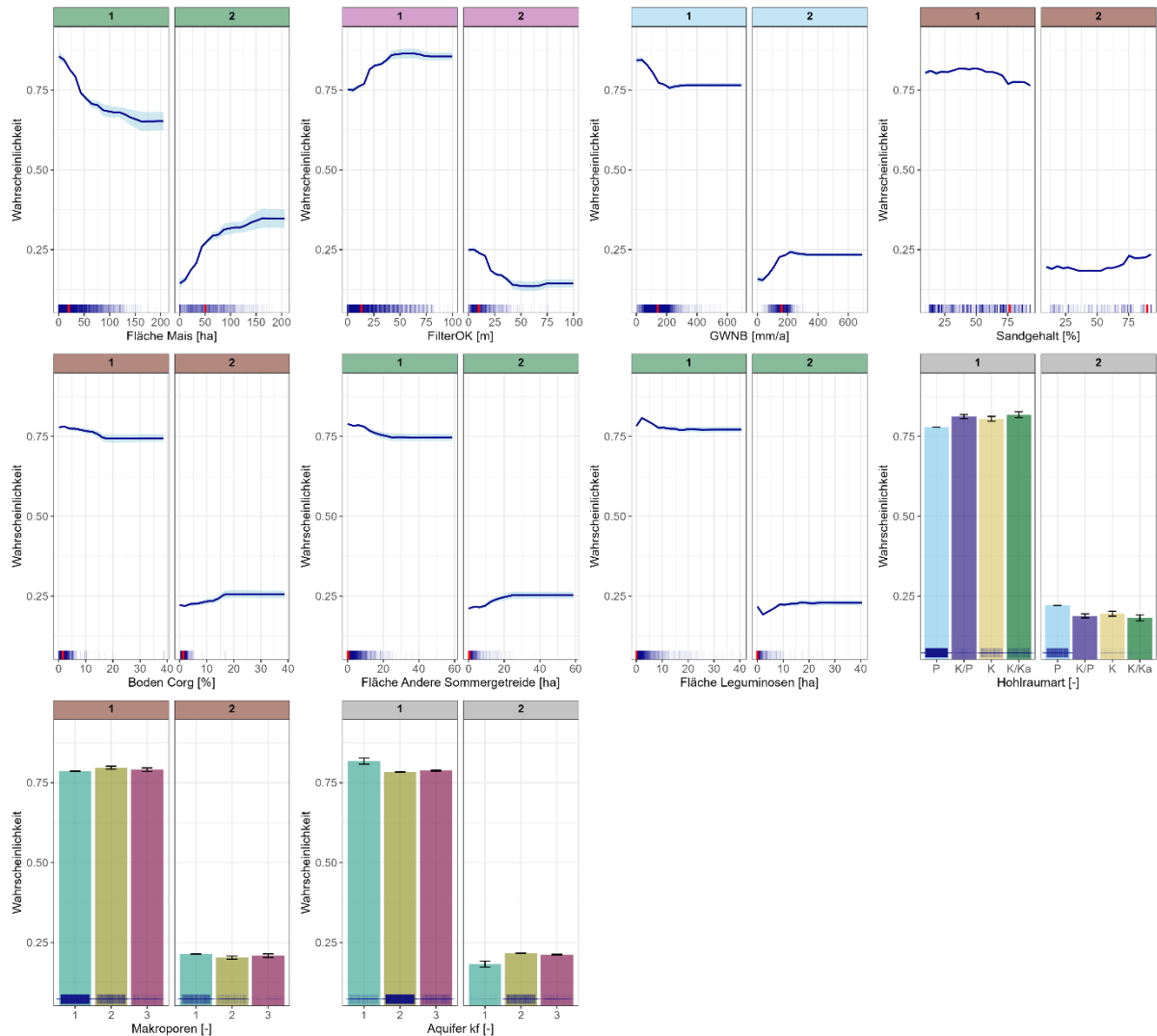
Die Feature-Effekte von Metolachlor OXA (Abbildung 43) weichen nicht nennenswert von den Effekten von Metolachlor ESA ab. Auch der Metabolit NOA 413173 (Abbildung 44) verhält sich ähnlich. Abweichungen sind nur bei den Makroporen zu sehen, die hier kaum Effekt zeigen.

Abbildung 43: Feature-Effekte von Metolachlor OXA im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.



Quelle: eigene Darstellung BGR.

Abbildung 44: Feature-Effekte von Metolachlor NOA 413173 im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.



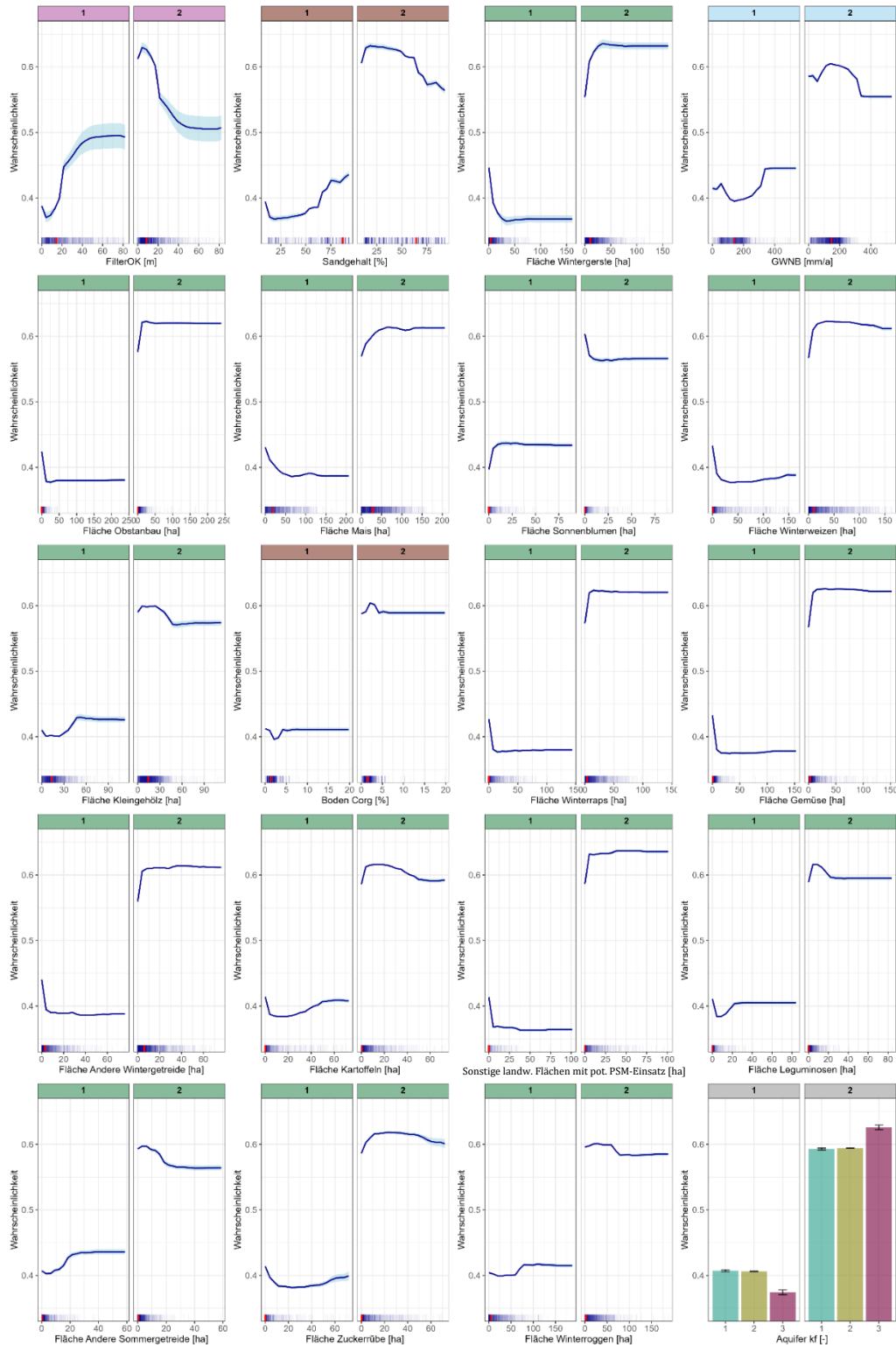
Quelle: eigene Darstellung BGR.

Trifluoressigsäure

Die Tiefe der Filteroberkante, das wichtigste Feature im TFA-Modell (Abbildung 45), zeigt die stärksten Effektamplituden und nach einem anfänglichen leichten Anstieg einen generell abfallenden Trend der Wahrscheinlichkeit der Konzentrationsklasse 2 bei zunehmender Tiefe. Nach einem anfänglichen Anstieg zeigt auch der Sandgehalt im Boden einen fallenden Trend für Klasse 2 mit zunehmendem Sandgehalt ab etwa 50%. Die Anbaufläche an Wintergerste im Puffer zeigt für einen Wertebereich bis etwa 50 ha einen starken Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2, danach flacht die Kurve ab. Alle nachfolgenden Kultur-Features zeigen einen ähnlichen, wenn auch z.T. steileren und begrenzteren Effekt. Der Effekt scheint darauf zu beruhen, ob die Kultur im Puffer vorkommt oder nicht, während höhere Flächenanteile dabei nicht ins Gewicht fallen. Lediglich das Feature Anbaufläche an Sonnenblumen zeigt einen gegensätzlichen Zusammenhang. Die Grundwasserneubildung weist über den gesamten Wertebereich einen leichten Abfall der Wahrscheinlichkeit für Konzentrationsklasse 2, mit je nach Wertebereich auch gegensätzlicher Dynamik. Der organische Kohlenstoffgehalt zeigt sehr

geringe Effekte. Beim Aquifer-kf zeigt sich eine erhöhte Wahrscheinlichkeit der Konzentrationsklasse bei Aquifer-kf-Klasse 3.

Abbildung 45: Feature-Effekte der Trifluoressigsäure im binären Random-Forest Modell der reduzierten Feature-Auswahl.



Quelle: eigene Darstellung BGR.

Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt. veranschaulichen die Seeds. Zur besseren Übersicht werden nicht alle Features gezeigt.

Vergleich der Feature-Effekte zwischen Random-Forest und Neuronalem Netz

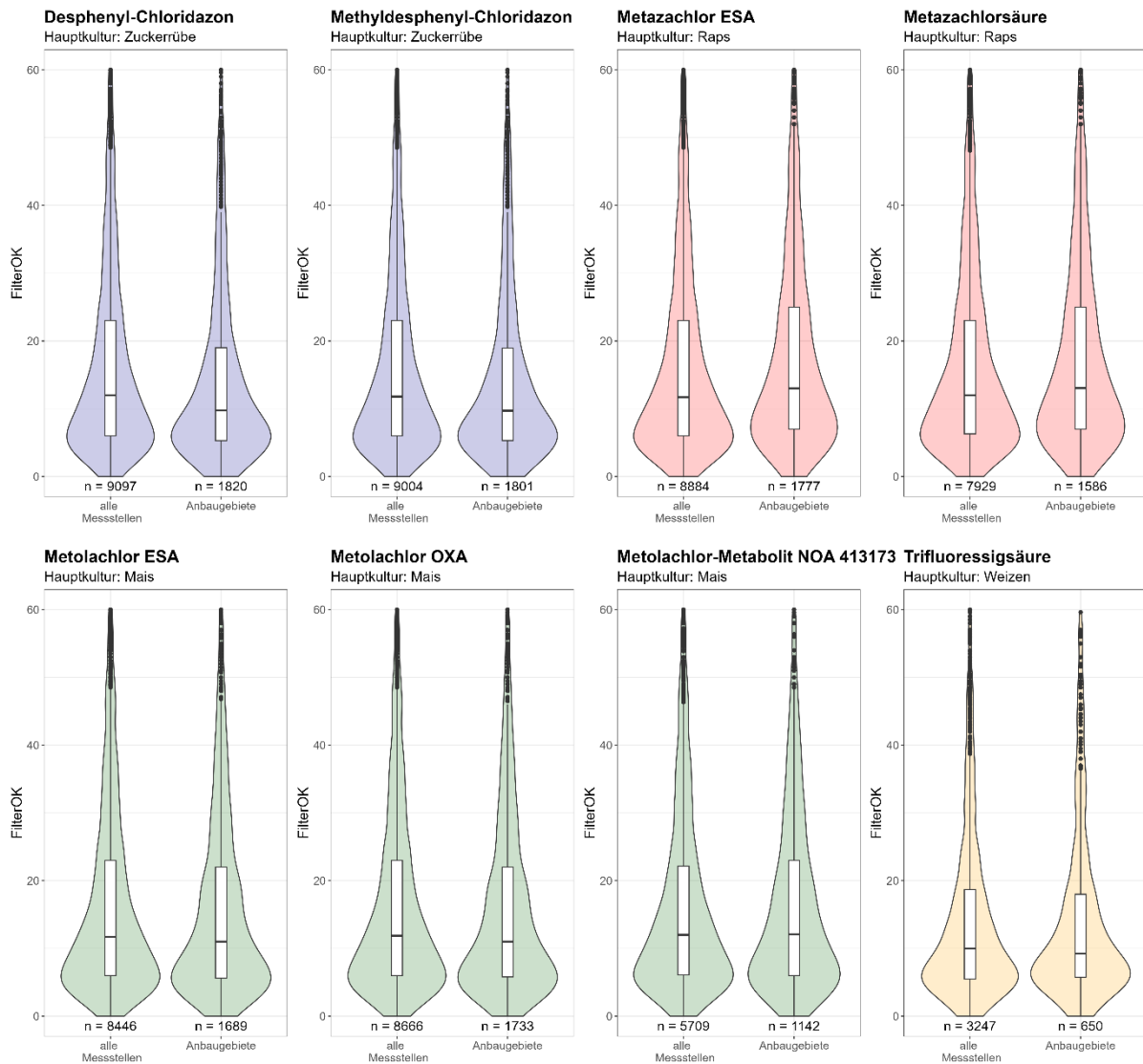
Die Partial-Dependence-Plots der Neuronalen Netze befinden sich im Anhang A.7. Ein qualitativer Abgleich der Modelle hinsichtlich der Feature-Effekte zeigt viel Übereinstimmung. Bei den Chloridazon-Metaboliten sind grundsätzlich vergleichbare Effekte zu sehen, einzelne Effektkurven (organischer Kohlenstoff) verlaufen beim NN-Modell flacher. Unterschiede werden bei Metazachlor ESA deutlich, da die NN-Modelle bei der Grundwasserneubildung und Sandgehalt keinen starken Effekt identifizieren. Bei der Metazachlorsäure finden sich größtenteils Übereinstimmungen. Bei Metolachlor ESA identifiziert das NN-Modell einen negativen Zusammenhang zwischen ansteigendem organischen Kohlenstoffgehalt und der Wahrscheinlichkeit der Konzentrationsklasse 2. Gleichzeitig wird bei der Grundwasserneubildung relativ viel Variabilität zwischen den einzelnen Teildatensätze sichtbar. Bei Metolachlor OXA zeigt sich beim NN-Modell eine stärker abfallende Effektkurve beim organischen Kohlenstoff, bei der Grundwasserneubildung ein stärkerer Anstieg. Bei Metolachlor NOA 413173 wird bei Sand kein starker Effekt gefunden, beim organischen Kohlenstoff findet sich relativ viel Variation zwischen den Teildatensätzen. Bei TFA ist der stärkere negative Zusammenhang zwischen organischem Kohlenstoff und der Konzentrationsklasse auffällig, der im RF-Modell nicht so stark ausgeprägt ist. Insgesamt lässt sich beobachten, dass die Effektkurven bei den NN-Modellen qualitativ den der RF-Modelle ähneln und die Modelle grundsätzlich vergleichbare Zusammenhänge aus den Features entnehmen. Gleichzeitig verlaufen die Kurven der NN-Modelle deutlich glatter. Es ist daher von Vorteil, die Effektkurven der Algorithmen für die Interpretation der Ergebnisse gemeinsam zu diskutieren. Starke Effekte der wichtigsten Features (Hauptkulturen, Filteroberkante) stimmen zwischen den Modellen weitestgehend überein. Unterschiede finden sich bei Features mit schwächeren Effekten, u.a. dem organischen Kohlenstoff im Boden.

3.2.1.4 Einfluss der Standortfaktoren in Anbaugebieten

Um das Verständnis dessen, was die Modelle aus den Features lernen, weiter zu vertiefen, wurde die Verteilung der Feature-Werte an Messstellen in den Anbaugebieten mit allen Messstellen verglichen. Messstellen, in deren Puffer die Anbaufläche der Hauptkultur für einen Stoff das 80. Perzentil der Verteilung der Anbaufläche an allen Messstellen darstellt oder übersteigt, wurden dabei als „Anbaugebiete“ definiert. Die Tiefe der Filteroberkante unterscheidet sich für keinen der acht Metaboliten signifikant zwischen allen Messstellen und den Messstellen in den Anbaugebieten (Abbildung 46). Der Median des Sandgehalts ist bei den Chloridazon- und Metazachlor-Metaboliten sowie bei TFA geringer als bei allen Messstellen (Abbildung 47). Bei den Metolachlor-Metaboliten zeigen sich kaum Unterschiede zwischen den Medianen. Der Median der Grundwasserneubildung ist bei den Chloridazon-Metaboliten niedriger in den Anbaugebieten, bei den Metazachlor-Metaboliten und bei TFA ist dies noch deutlicher ausgebildet. Die Metolachlor-Metaboliten weisen jedoch höhere Mediane in den Anbaugebieten auf (Abbildung 48). Während bei den Chloridazon-Metaboliten keine nennenswerten Unterschiede der Verteilung der Messstellen in den vier Hohlraumarten zwischen allen Messstellen und Messstellen in Anbaugebieten zu verzeichnen sind, kommt es bei den Metazachlor-Metaboliten und bei TFA zu einem höheren Anteil an Messstellen in Kluft- bzw. Kluft/Karst-Grundwasserleitern in den Anbaugebieten (Abbildung 49). Bei den Metolachlor-Metaboliten zeigt sich wiederum ein gegensätzlicher Zusammenhang, da in den Anbaugebieten kaum noch Messstellen an Kluft- oder Karstaquiferen vorhanden sind. Über alle Messstellen gesehen dominiert bei allen acht Abbauprodukten die Makroporenklasse 1 (kein Potential für Makroporenbildung). In den Anbaugebieten kommt es bei den Metolachlor-Metaboliten noch zu einer Verstärkung dieser Tendenz. Bei allen anderen Metaboliten kommt es in den Anbaugebieten zu einer klaren Verschiebung zugunsten der Makroporenklasse 2, im Fall der

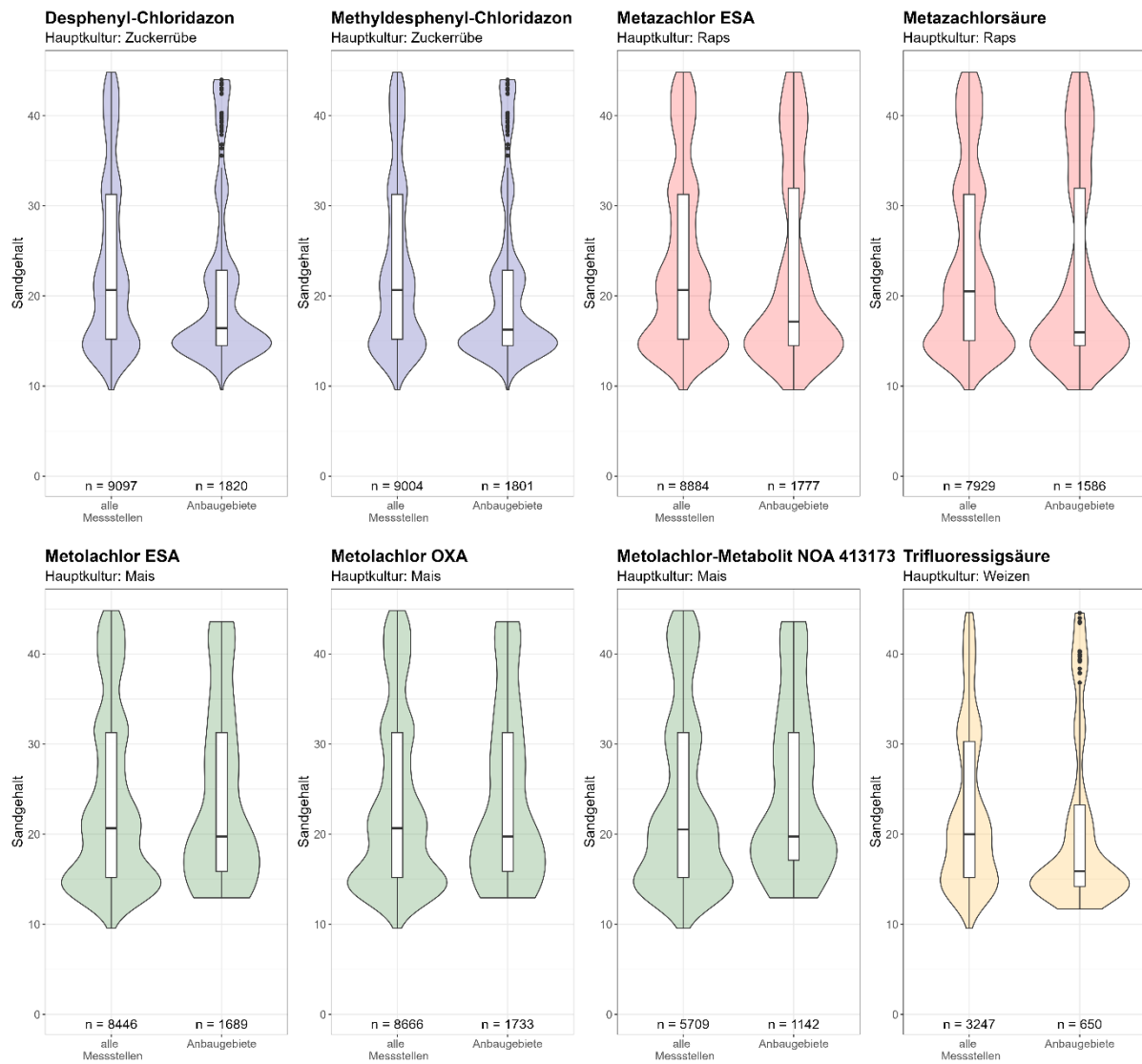
Chloridazon-Metaboliten auch zu einer Verschiebung zugunsten der Makroporenklasse 3, allerdings in geringerem Umfang (Abbildung 50). Die Verteilung des organischen Kohlenstoffs zeigte kaum Unterschiede zwischen allen Messstellen und den Anbaugebieten (Abbildung 51). Bei den Chloridazon-Metaboliten sind nur geringfügige Verschiebungen der hydraulischen Leitfähigkeit (Aquifer-kf-Klassen) zu beobachten, bei den Metazachlor-Metaboliten und TFA zeigt sich eine Verschiebung hin zu anteilig mehr Messstellen in Klasse 1 und weniger in Klasse 3. Bei den Metolachlor-Metaboliten ist ein umgekehrter Trend zu verzeichnen (Abbildung 52).

Abbildung 46: Verteilung der Werte der Filteroberkante in den Anbaugebieten.



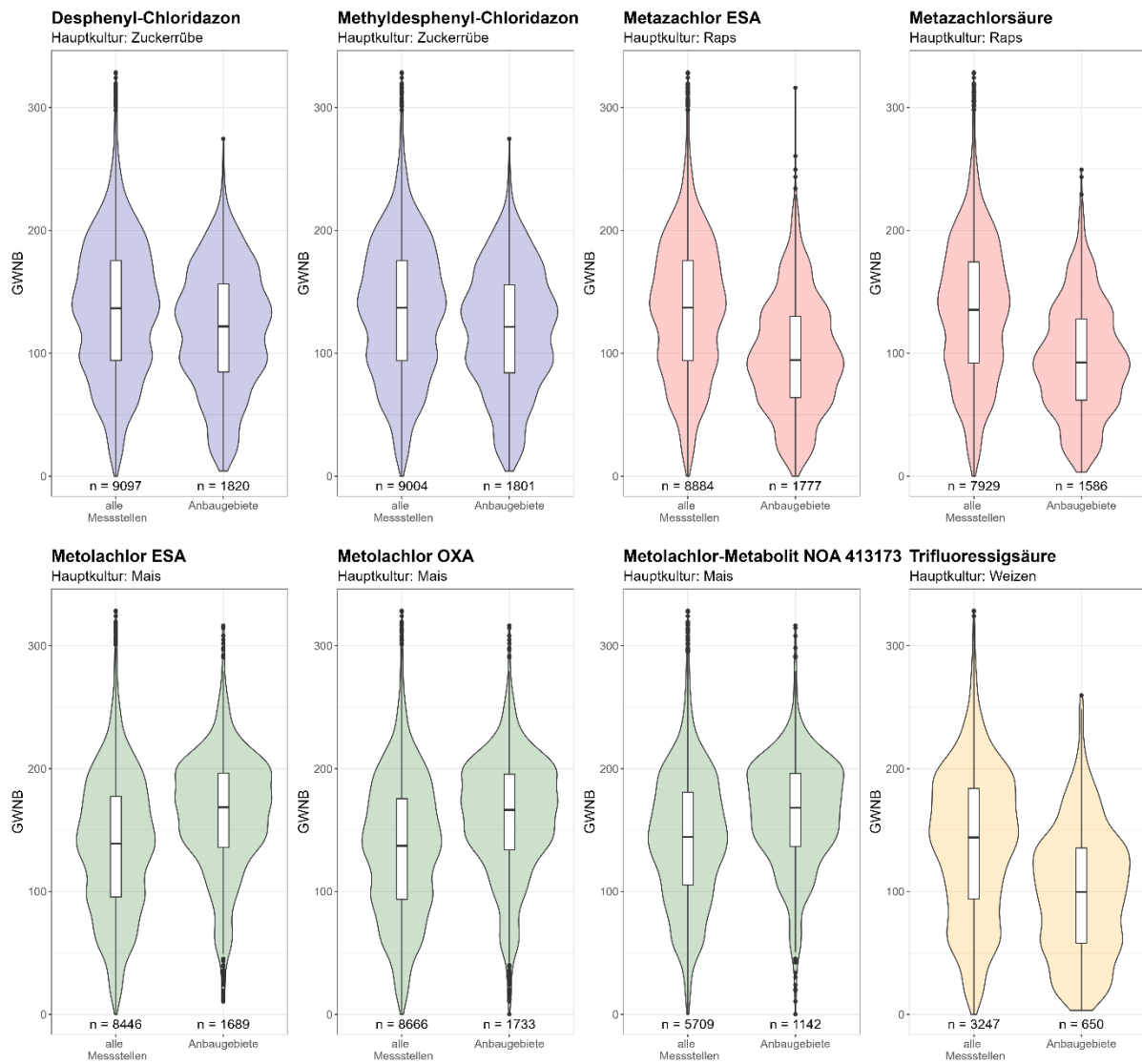
Quelle: eigene Darstellung BGR.

Abbildung 47: Verteilung der Werte des Sandgehalts in den Anbaugebieten.



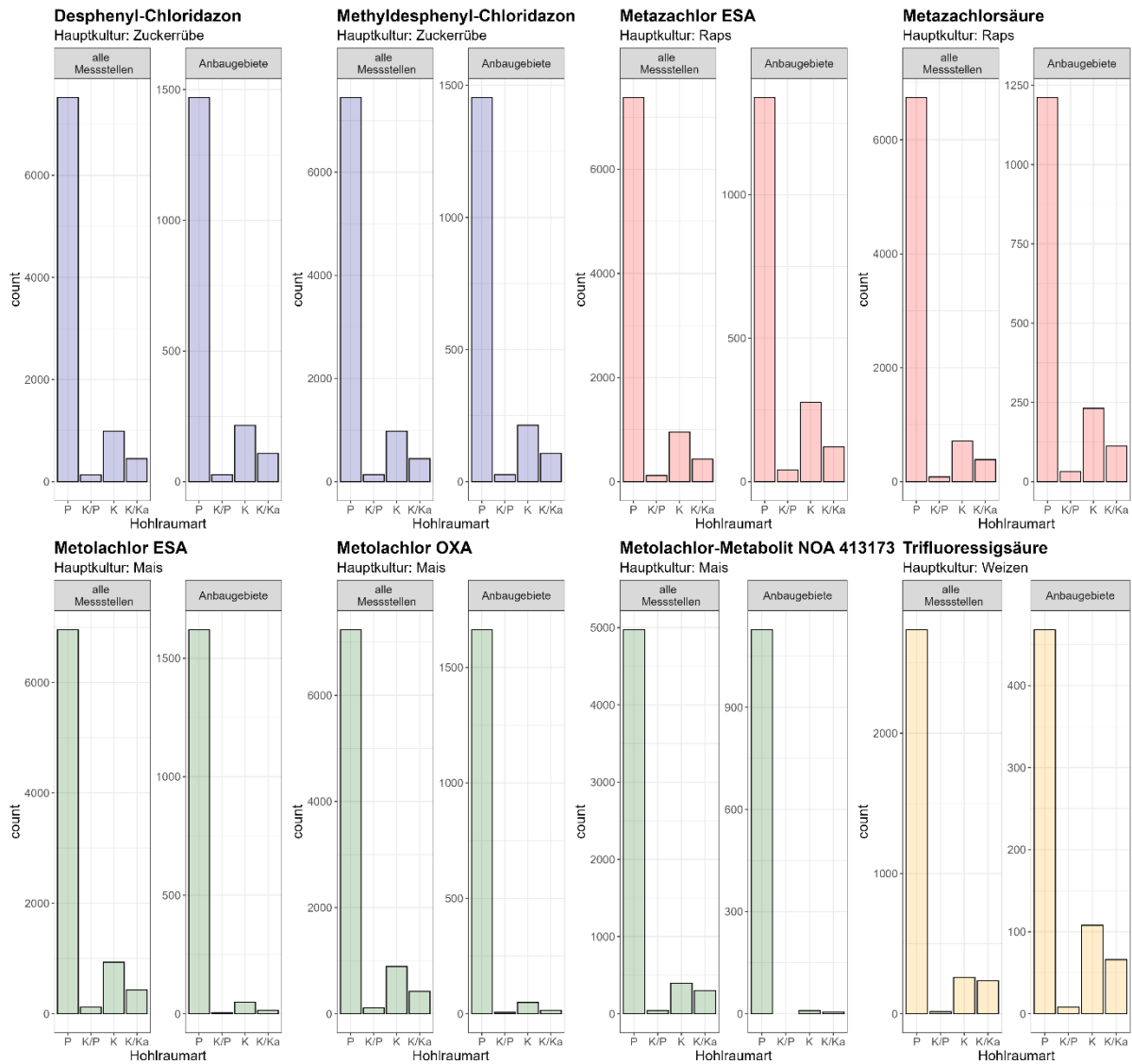
Quelle: eigene Darstellung BGR.

Abbildung 48: Verteilung der Werte der Grundwasserneubildung in den Anbaugebieten.



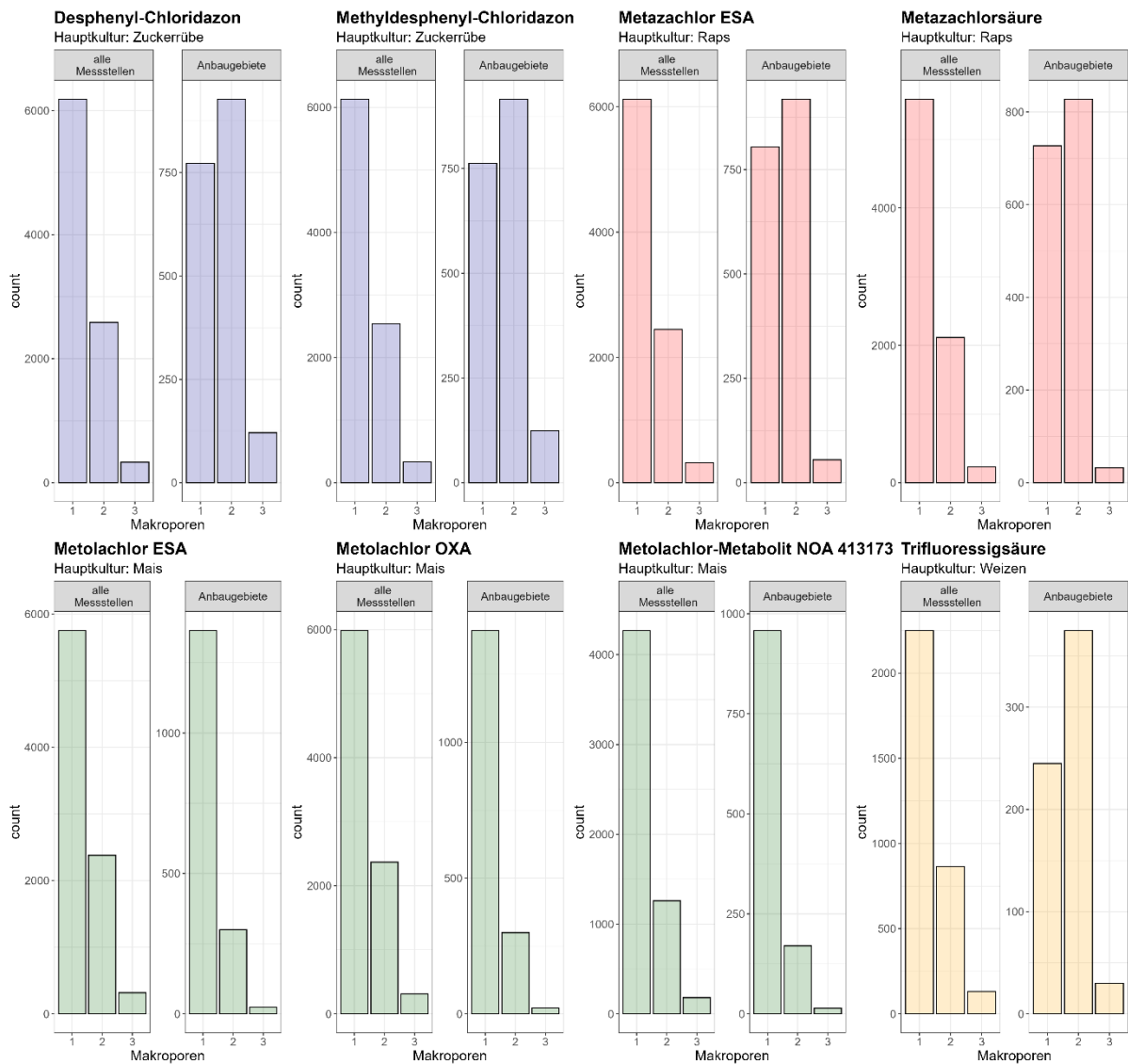
Quelle: eigene Darstellung BGR.

Abbildung 49: Verteilung der Kategorien der Hohlräumart in den Anbaugebieten.



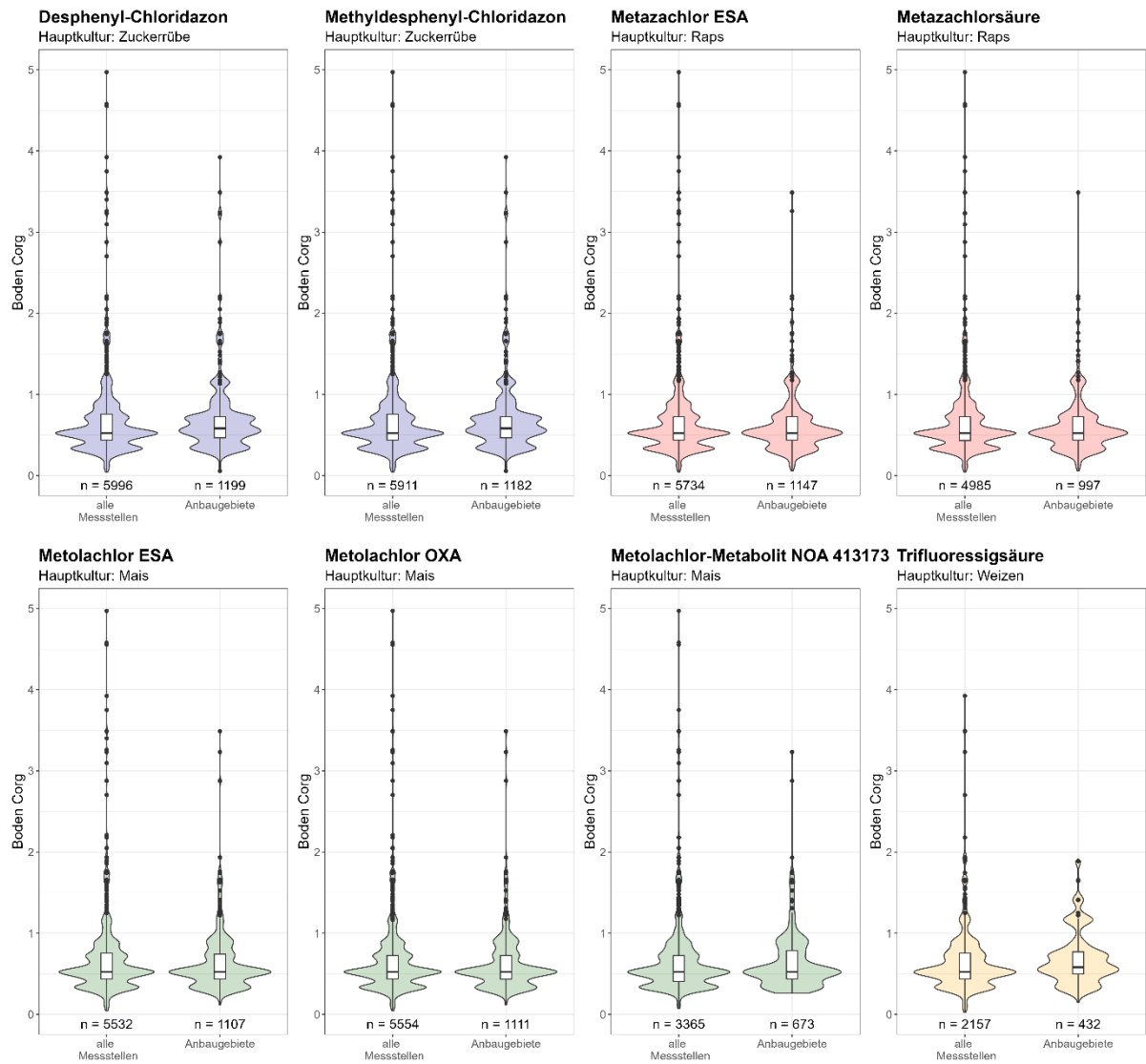
Quelle: eigene Darstellung BGR.

Abbildung 50: Verteilung der Werte der Kategorien der Makroporen in den Anbaubereichen.



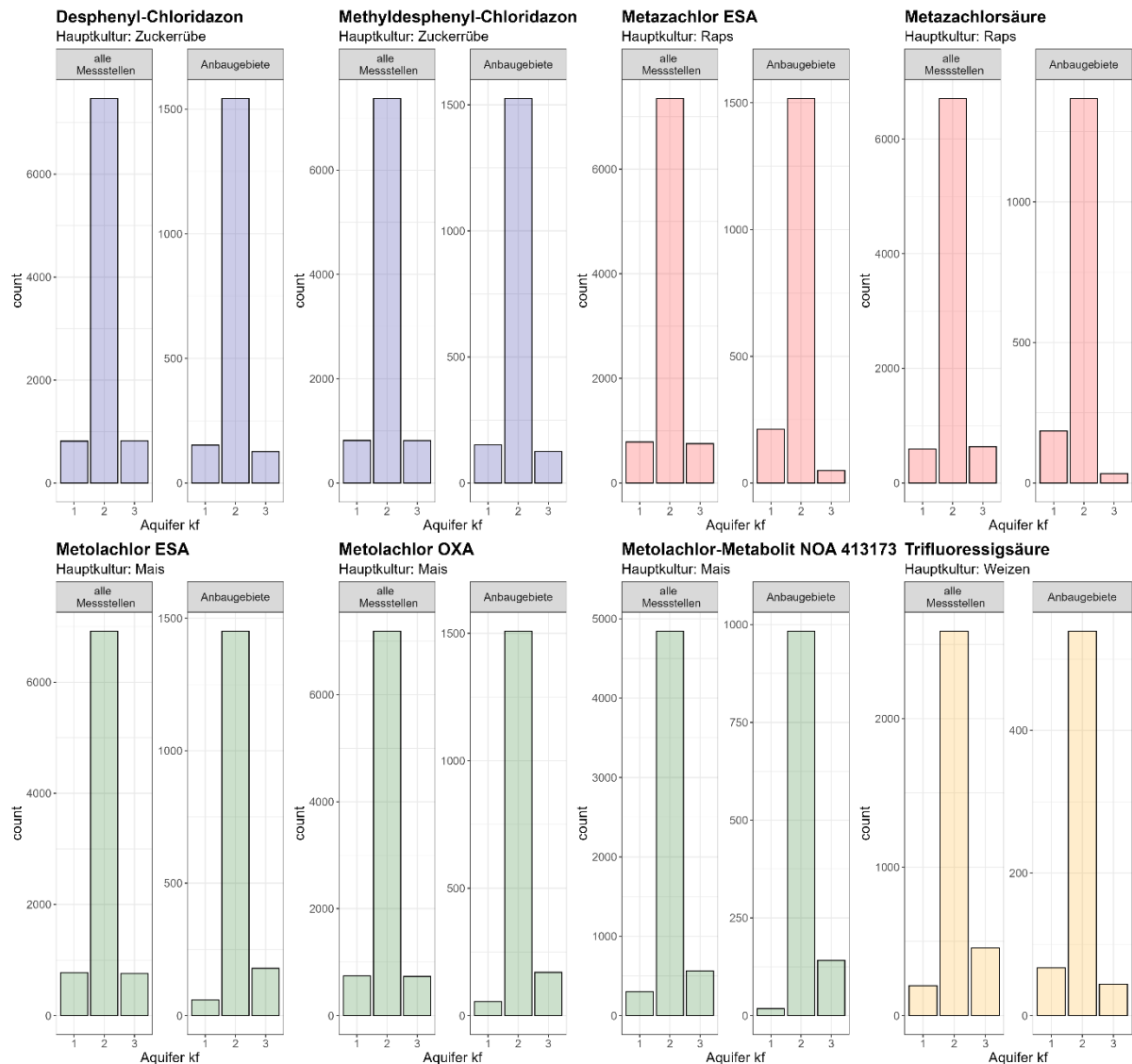
Quelle: eigene Darstellung BGR.

Abbildung 51: Verteilung der Werte des organischen Kohlenstoffs im Boden in den Anbaugebieten.



Quelle: eigene Darstellung BGR.

Abbildung 52: Verteilung der Werte der Kategorien der Durchlässigkeit des Grundwasserleiters in den Anbaugebieten.

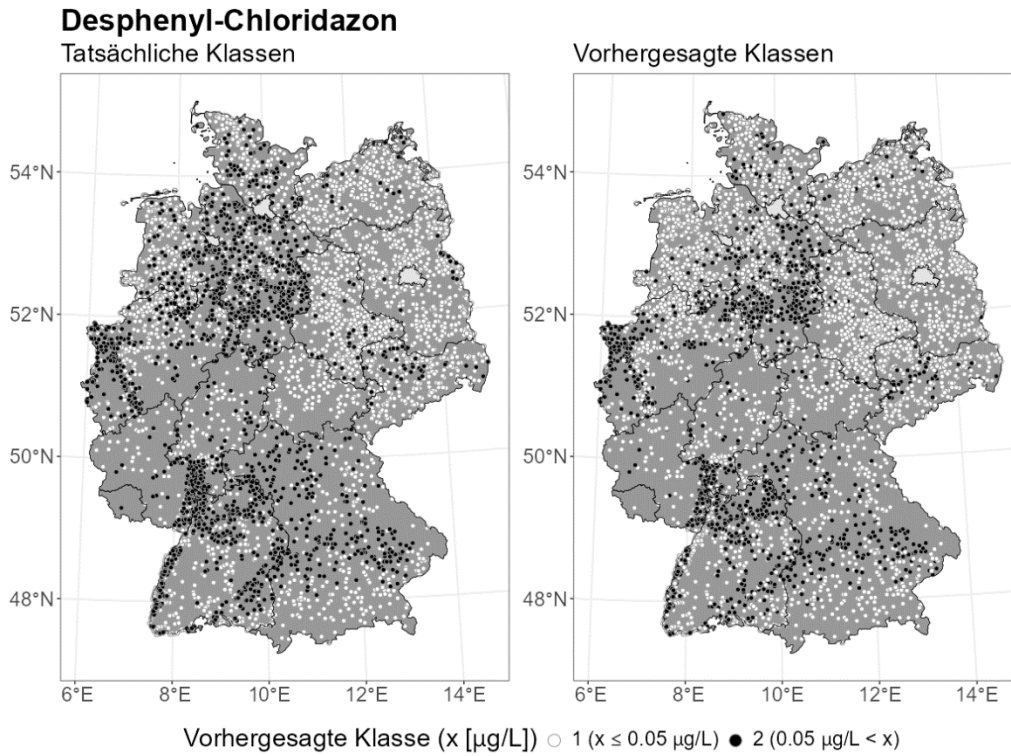


Quelle: eigene Darstellung BGR.

3.2.1.5 Räumliche Vorhersagen

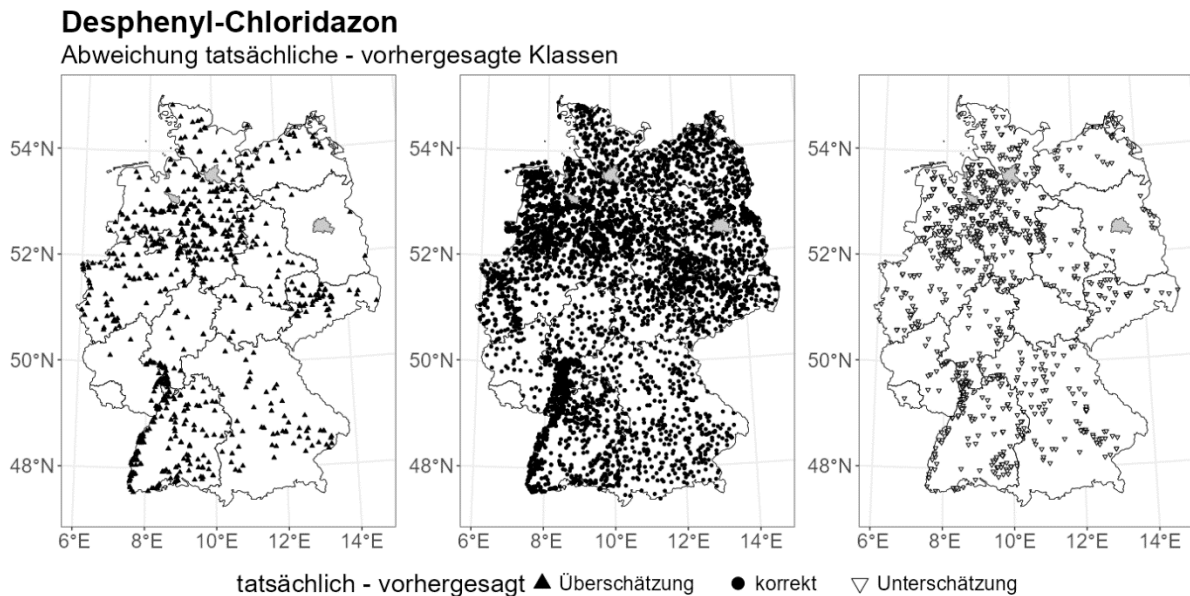
In den nachfolgenden Abbildungen finden sich die räumlichen Darstellungen der Messstellen, die über die Signatur jeweils nach tatsächlichen bzw. vorhergesagten Konzentrationsklassen pro Stoff gekennzeichnet sind.

Abbildung 53: Tatsächliche und vorhergesagte Konzentrationsklasse für Desphenyl-Chloridazon aller Grundwassermessstellen.



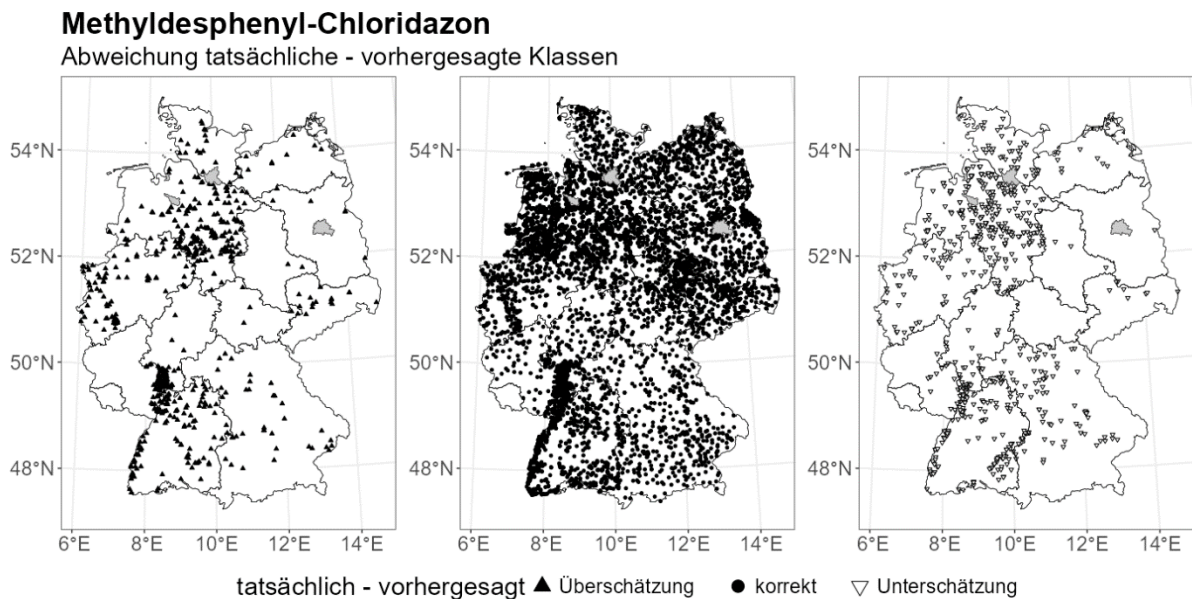
Quelle: eigene Darstellung BGR.

Abbildung 54: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Desphenyl-Chloridazon der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



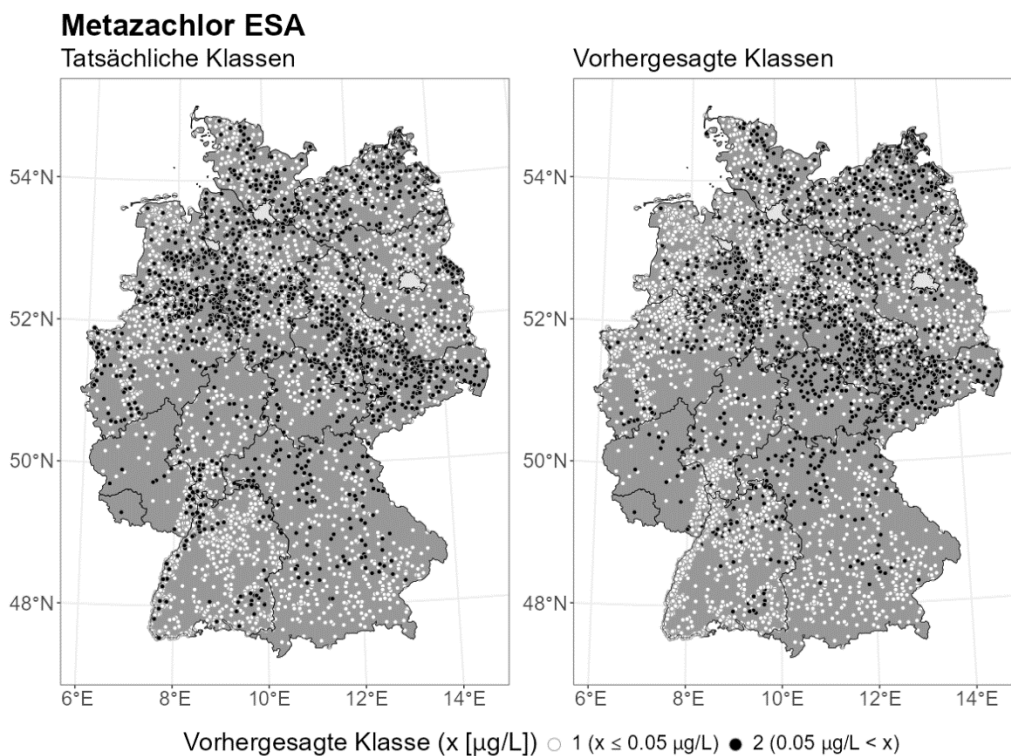
Quelle: eigene Darstellung BGR.

Abbildung 55: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Methyldesphenyl-Chloridazon der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



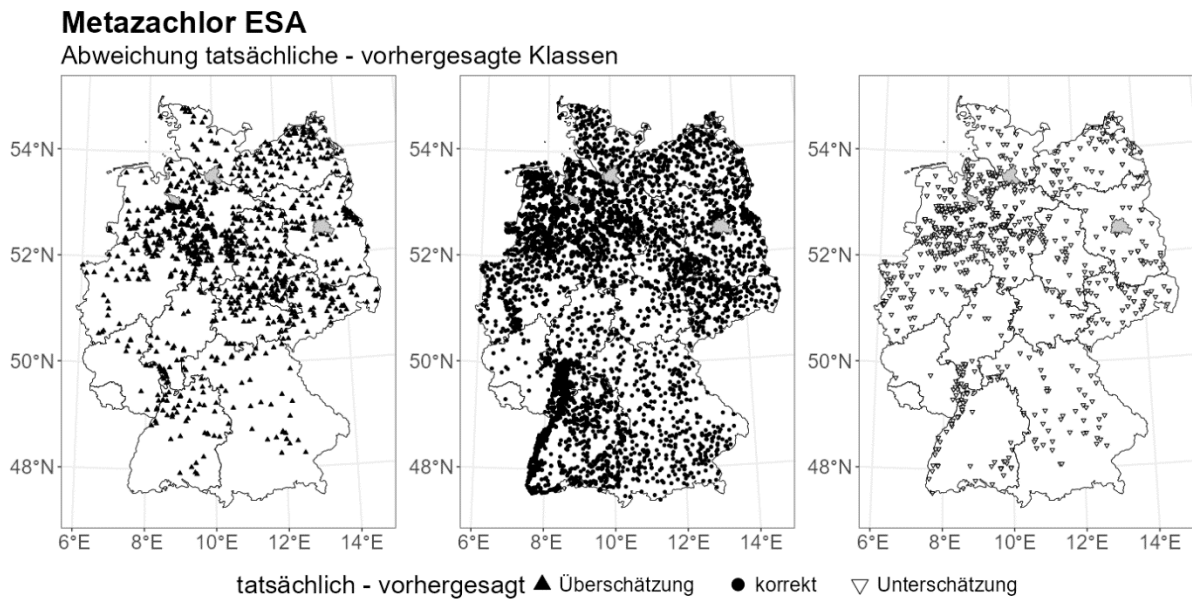
Quelle: eigene Darstellung BGR.

Abbildung 56: Tatsächliche und vorhergesagte Konzentrationsklasse für Metazachlor ESA aller Grundwassermessstellen.



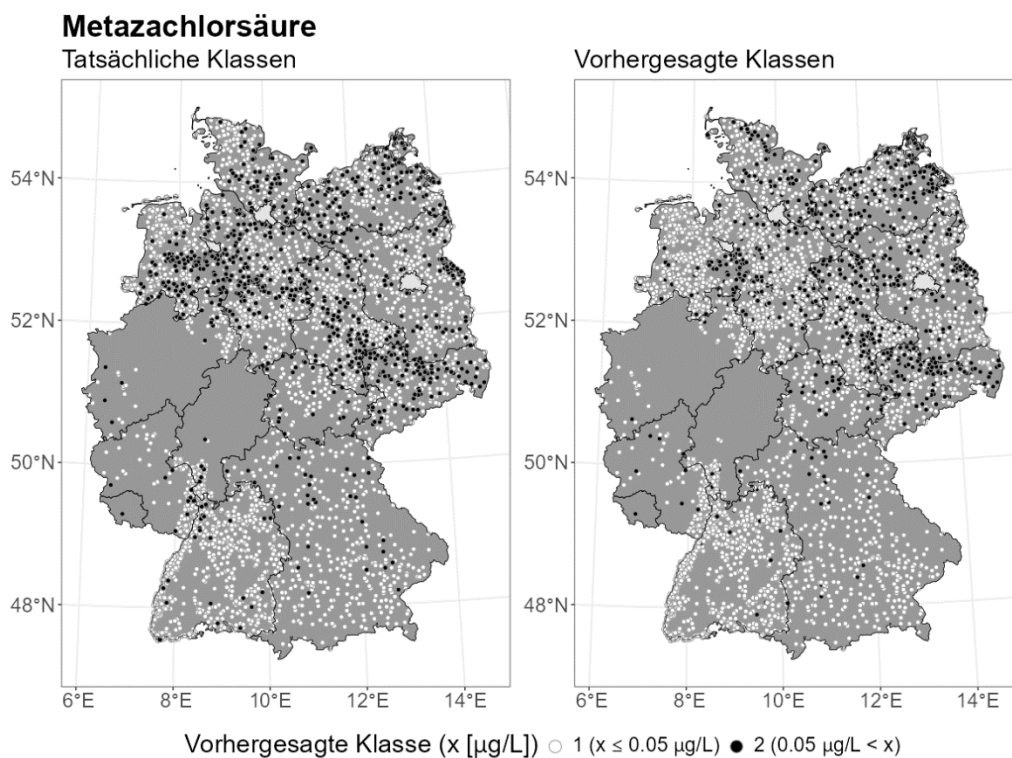
Quelle: eigene Darstellung BGR.

Abbildung 57: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metazachlor ESA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



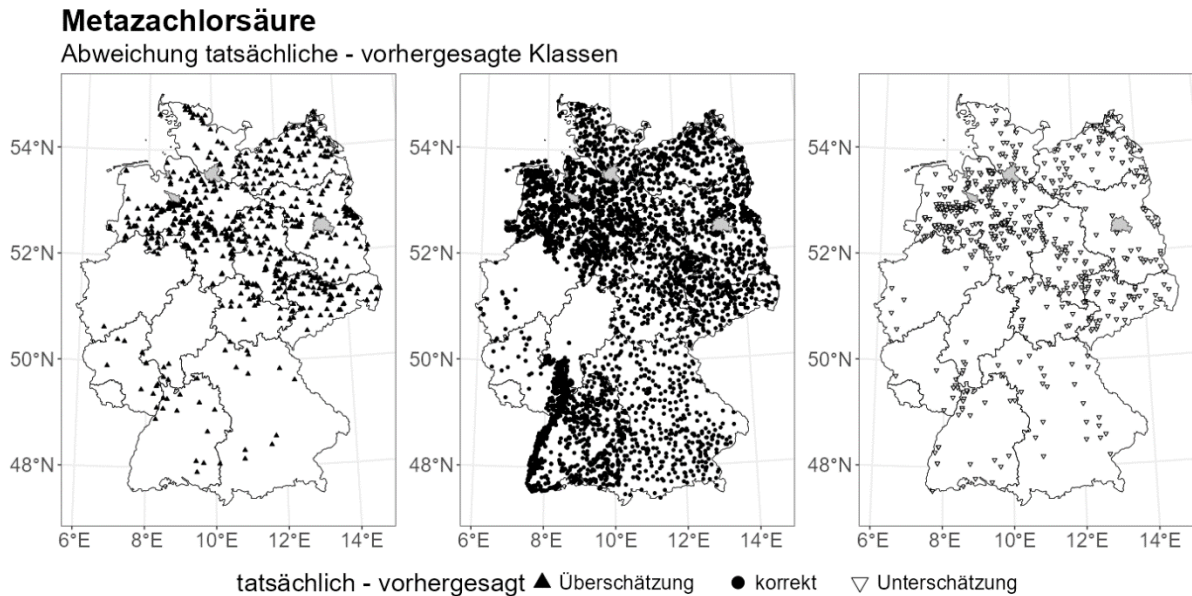
Quelle: eigene Darstellung BGR.

Abbildung 58: Tatsächliche und vorhergesagte Konzentrationsklasse für Metazachlorsäure aller Grundwassermessstellen.



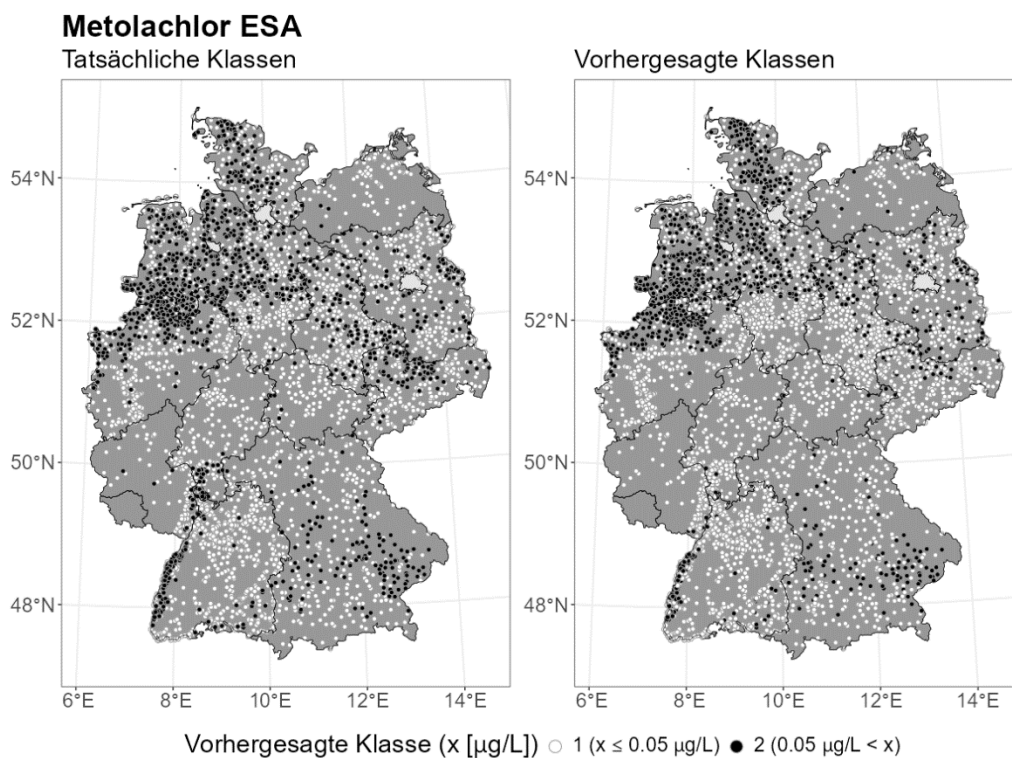
Quelle: eigene Darstellung BGR.

Abbildung 59: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metazachlorsäure der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



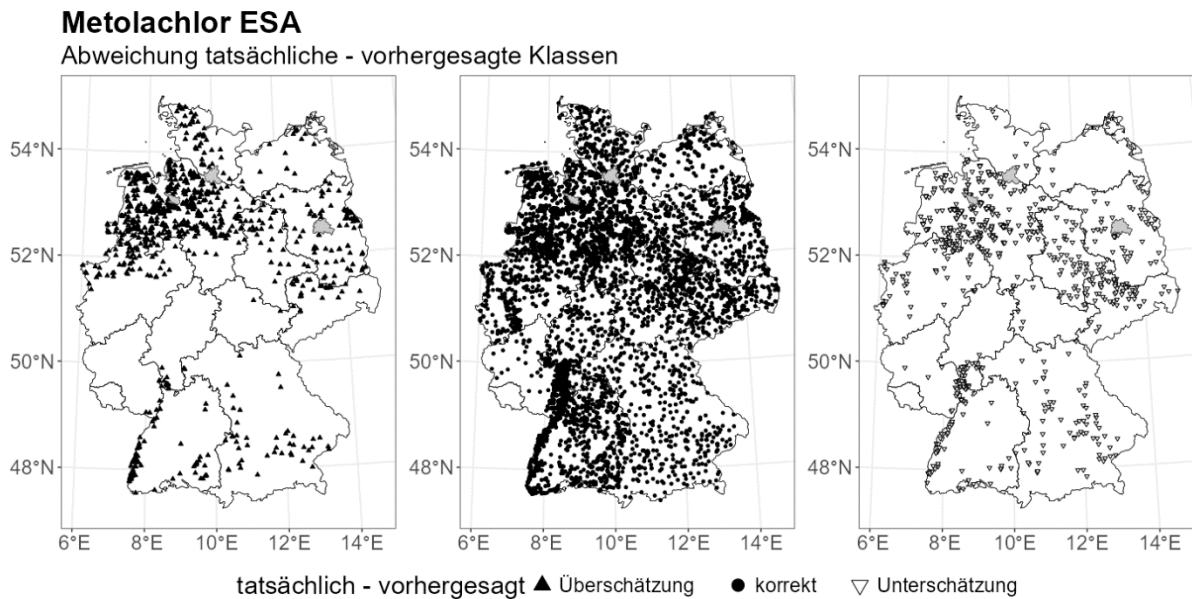
Quelle: eigene Darstellung BGR.

Abbildung 60: Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor ESA aller Grundwassermessstellen.



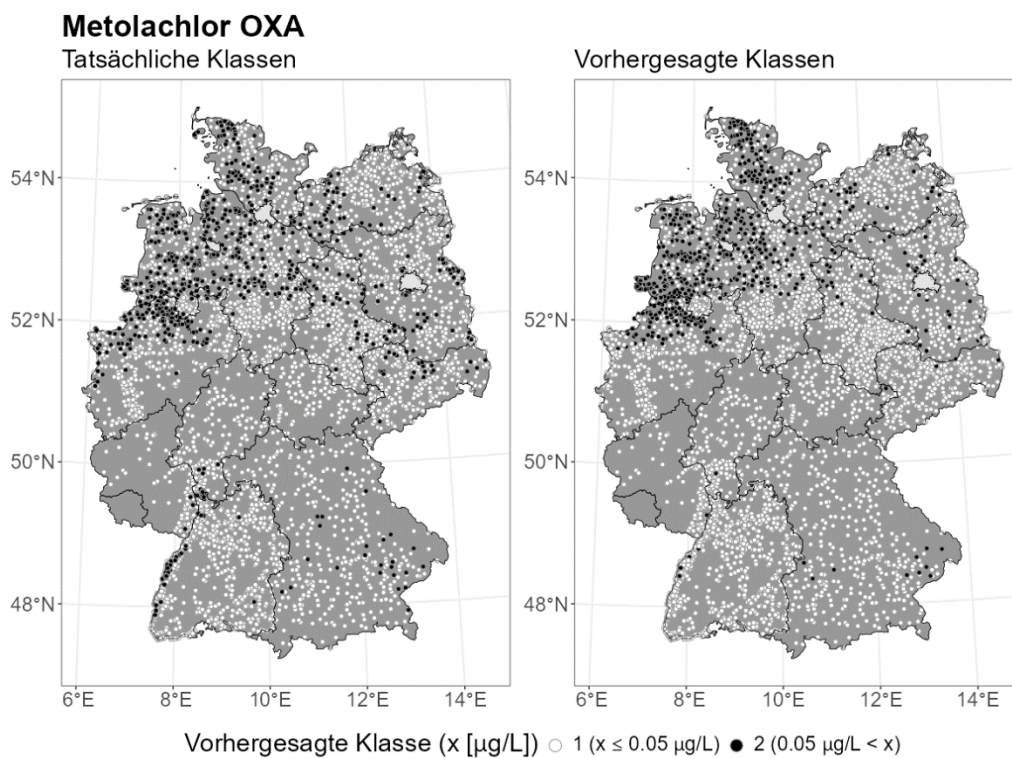
Quelle: eigene Darstellung BGR.

Abbildung 61: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor ESA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



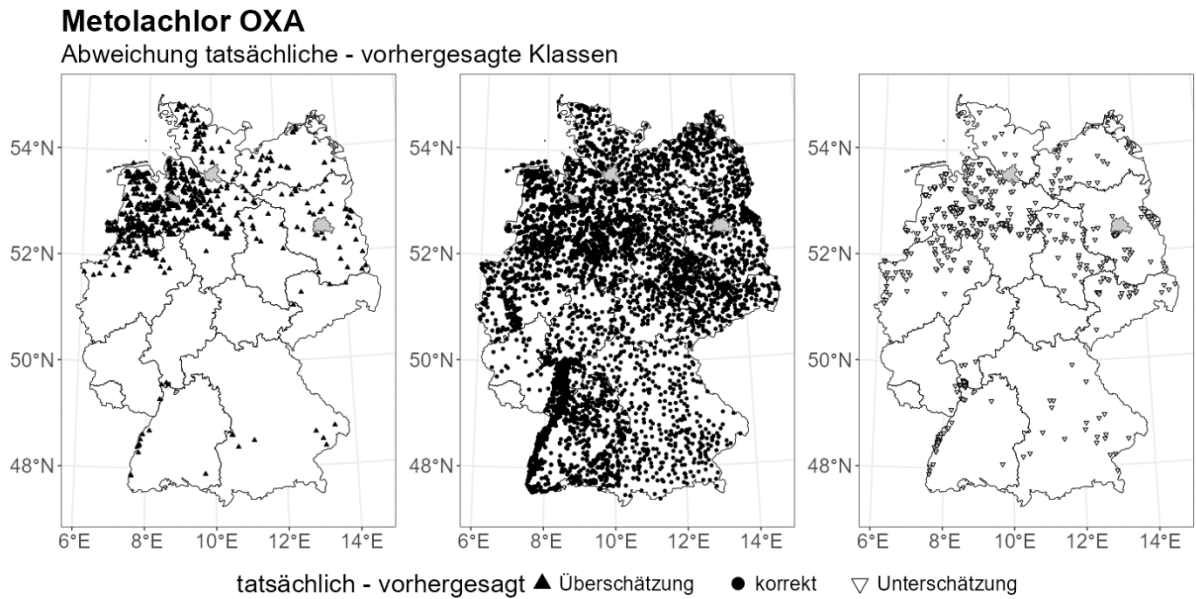
Quelle: eigene Darstellung BGR.

Abbildung 62: Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor OXA aller Grundwassermessstellen.



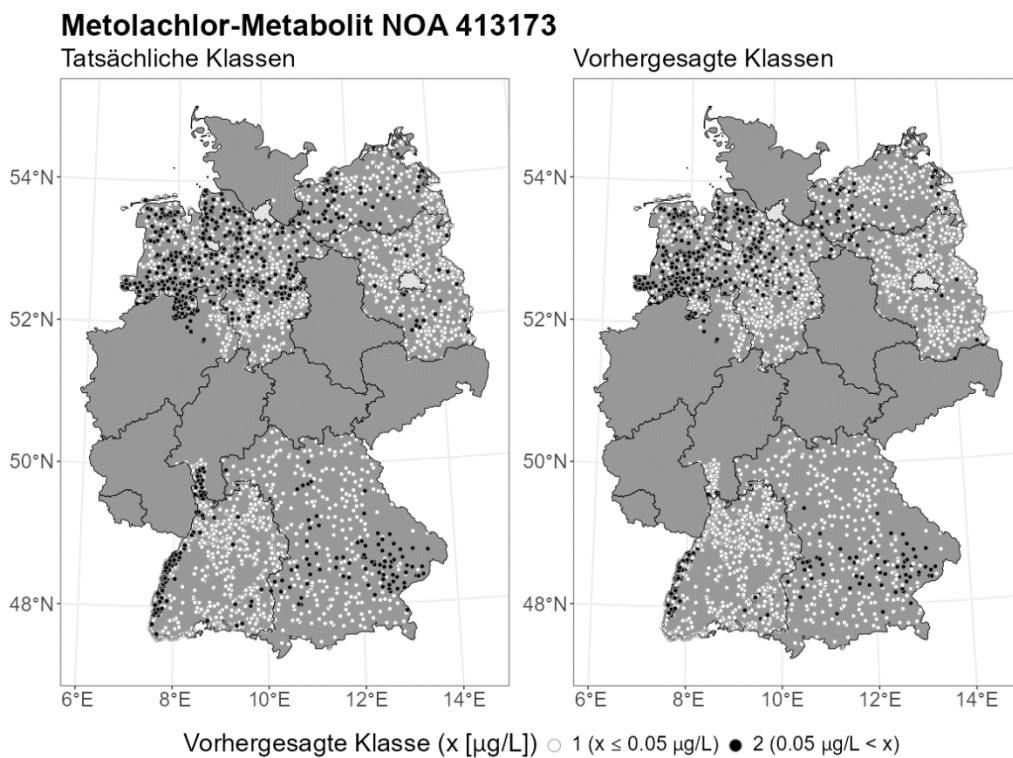
Quelle: eigene Darstellung BGR.

Abbildung 63: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor OXA der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



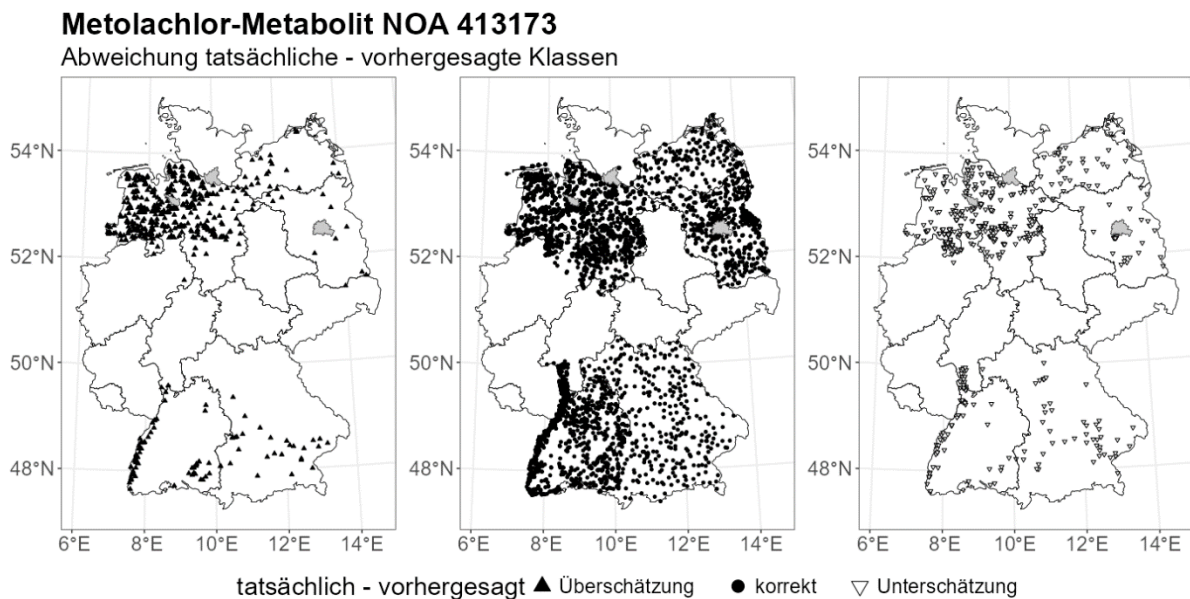
Quelle: eigene Darstellung BGR.

Abbildung 64: Tatsächliche und vorhergesagte Konzentrationsklasse für Metolachlor-Metabolit NOA 413173 aller Grundwassermessstellen.



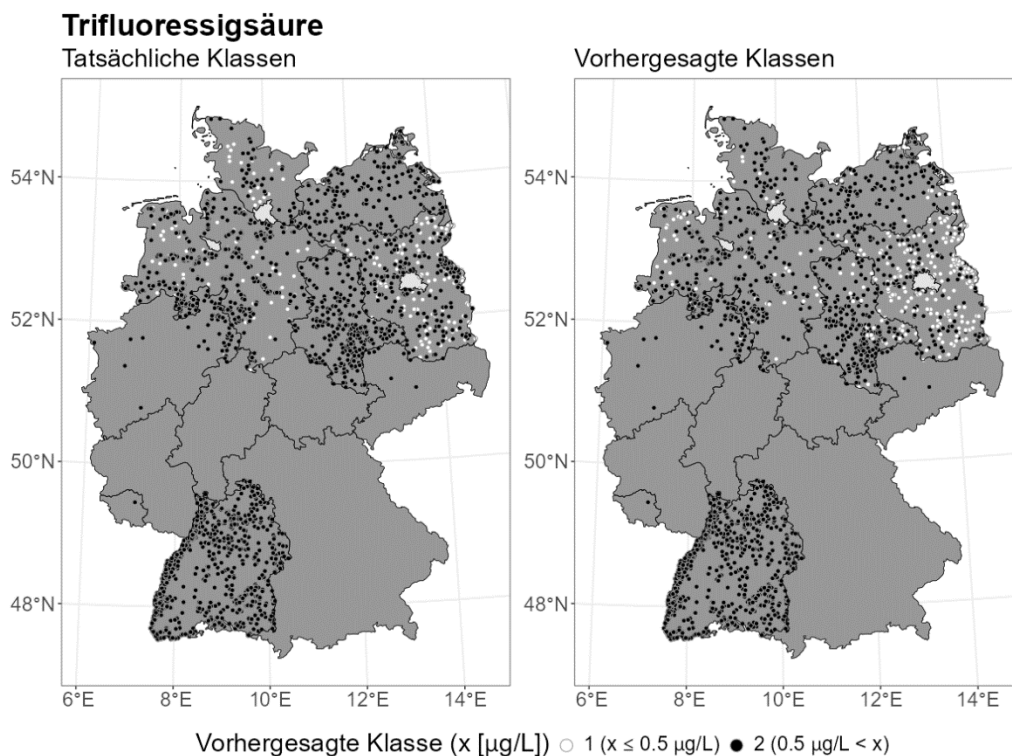
Quelle: eigene Darstellung BGR.

Abbildung 65: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Metolachlor-Metabolit NOA 413173 der Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



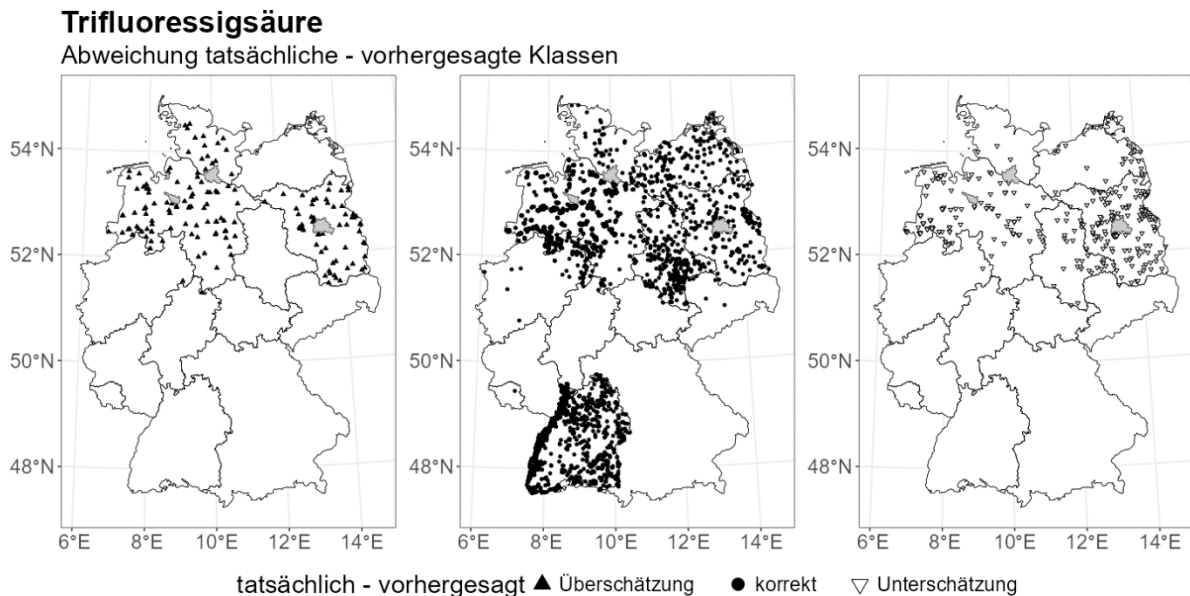
Quelle: eigene Darstellung BGR.

Abbildung 66: Tatsächliche und vorhergesagte Konzentrationsklasse für Trifluoressigsäure aller Grundwassermessstellen.



Quelle: eigene Darstellung BGR.

Abbildung 67: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen für Trifluoressigsäure der Konzentrationsklasse 2 ($> 0,5 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



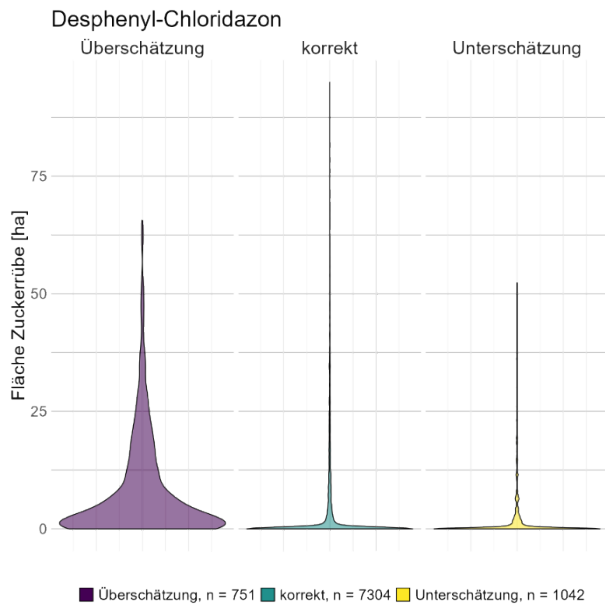
Quelle: eigene Darstellung BGR.

3.2.1.6 Analyse der Residuen

Eine Analyse der Residuen trägt dazu bei die von den Modellen erlernten Zusammenhänge genauer zu beleuchten. Mit Residuen sind die Messstellen gemeint, an denen es zu einer Abweichung zwischen der tatsächlichen und der vorhergesagten Konzentrationsklasse kommt. Es bietet sich an, die Charakteristika dieser Messstellen in Bezug auf die Featurewerte herauszuarbeiten. Dies dient einem besseren Verständnis, warum es zu Abweichungen gekommen sein könnte und wie die Modelle zukünftig verbessert werden können.

Messstellen, an denen es zu Überschätzungen kommt, weisen eine Verteilung mit etwas höheren Flächenangaben der Hauptkultur auf, während diese an Messstellen, an denen es zu Unterschätzungen kommt, etwas geringer ausfällt. Dies ist beispielhaft für Desphenyl-Chloridazon (Abbildung 68), Metazachlor ESA (Abbildung 69) und Metolachlor OXA (Abbildung 70) dargestellt. Gleichermäßen wurde auch die Verteilung der Werte der Filteroberkante untersucht. Messstellen, an denen es zu einer Überschätzung kommt, weisen oberflächennahe Filter auf, während Messstellen, an denen es zu Unterschätzungen kommt, oberflächenferne Filter aufweisen. Dies wird für Metolachlor NOA 413173 (Abbildung 71) und die Metazachlorsäure (Abbildung 72) gezeigt.

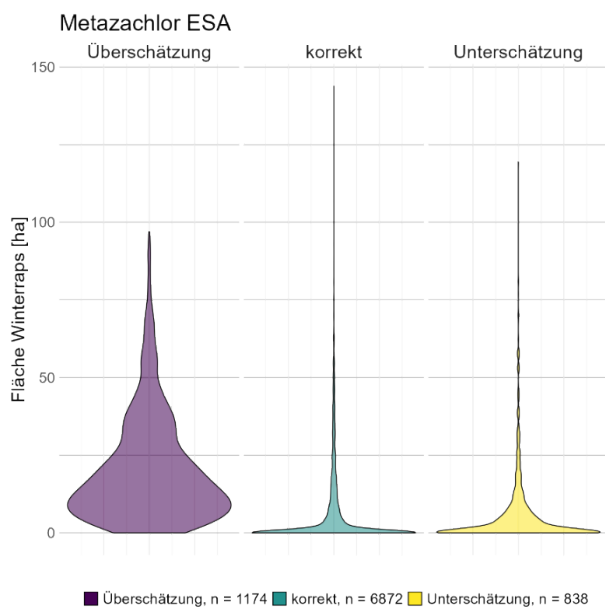
Abbildung 68: Desphenyl-Chloridazon: Die Verteilung der Anbaufläche der Hauptkultur (Zuckerrübe) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (mitte, grün) und Unterschätzung (rechts, gelb) als Violinenplot dargestellt.



Quelle: eigene Darstellung BGR.

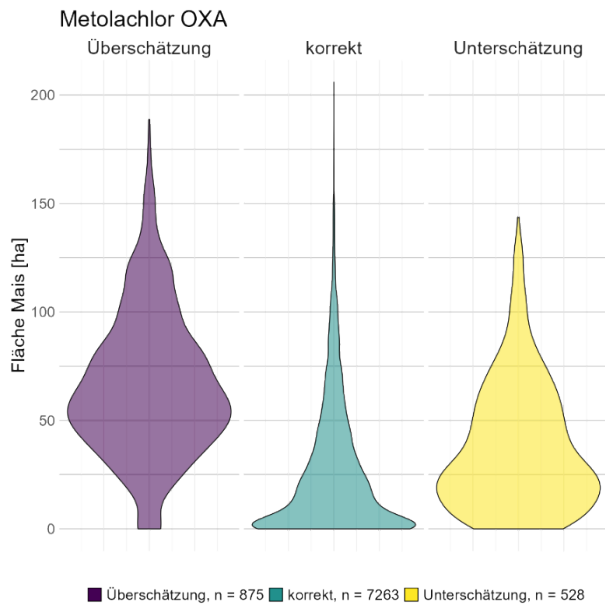
Die Anzahl der Messstellen ist jeweils in der Legende in Klammern gegeben. Binäre Random-Forest-Modelle, reduzierte Feature-Auswahl (hier und in den folgenden).

Abbildung 69: Metazachlor ESA: Die Verteilung der Anbaufläche der Hauptkultur (Winterraps) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Unterschätzung (rechts, gelb) als Violinenplot dargestellt.



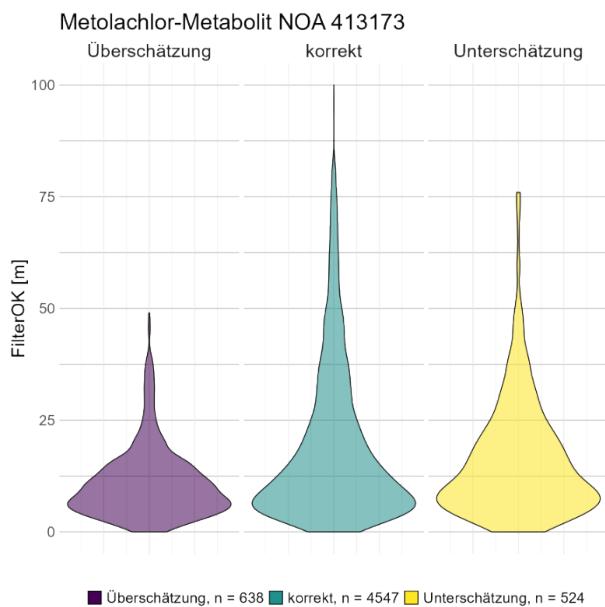
Quelle: eigene Darstellung BGR.

Abbildung 70: Metolachlor OXA: Die Verteilung der Anbaufläche der Hauptkultur (Mais) wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Unterschätzung (rechts, gelb) als Violinenplot dargestellt.



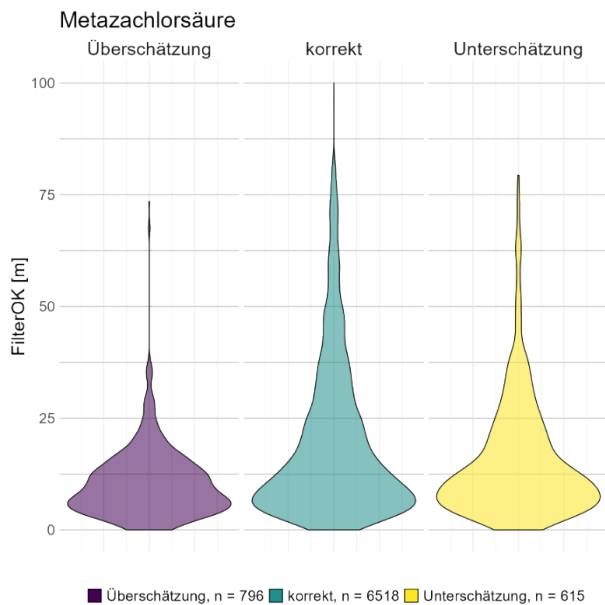
Quelle: eigene Darstellung BGR.

Abbildung 71: Metolachlor NOA 413173: Die Verteilung der Tiefe der Filteroberkante wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Unterschätzung (rechts, gelb) als Violinenplot dargestellt.



Quelle: eigene Darstellung BGR.

Abbildung 72: Metazachlorsäure: Die Verteilung der Tiefe der Filteroberkante wurde an den Messstellen der je drei Optionen (Überschätzung (links, lila), korrekte Vorhersage (Mitte, grün) und Unterschätzung (rechts, gelb) als Violinenplot dargestellt.



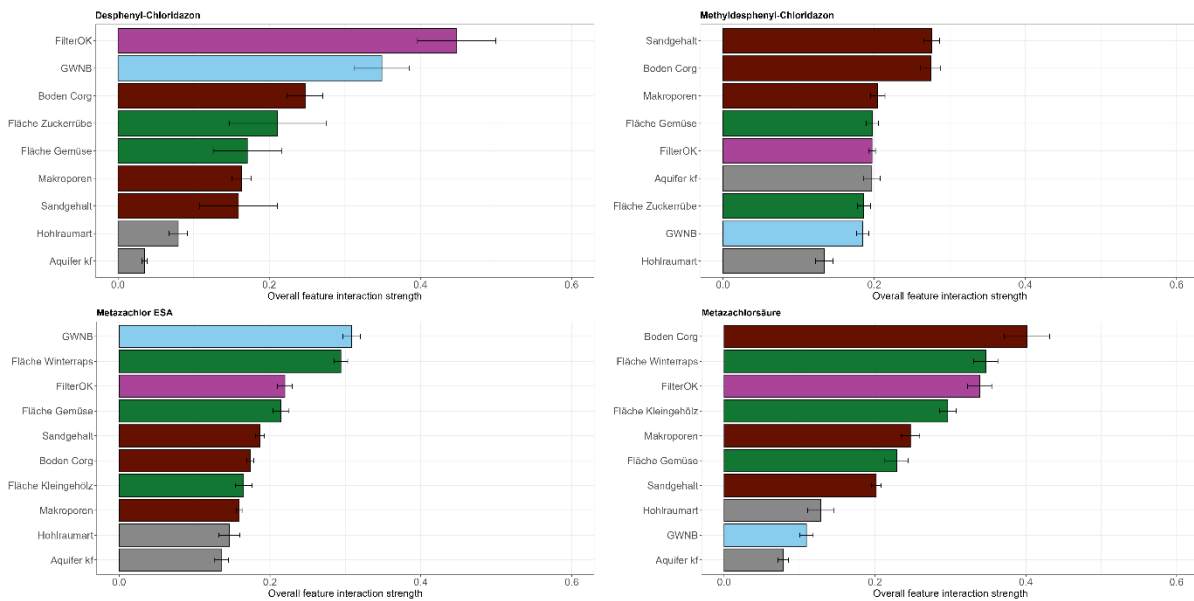
Quelle: eigene Darstellung BGR.

3.2.1.7 Feature-Interaktionen

Nachdem die relevanten Features identifiziert und ihre Effekte beleuchtet wurden, soll im Folgenden der Grad der Interaktion, d.h. der Wechselwirkung zwischen den Features und deren Einfluss auf die Vorhersage beleuchtet werden (siehe 2.2.3.7). Damit soll die Frage beantwortet werden, ob die Modelle nichtlineare oder eher additive Zusammenhänge zwischen den Features lernen, was zur Interpretation der Modelle beiträgt.

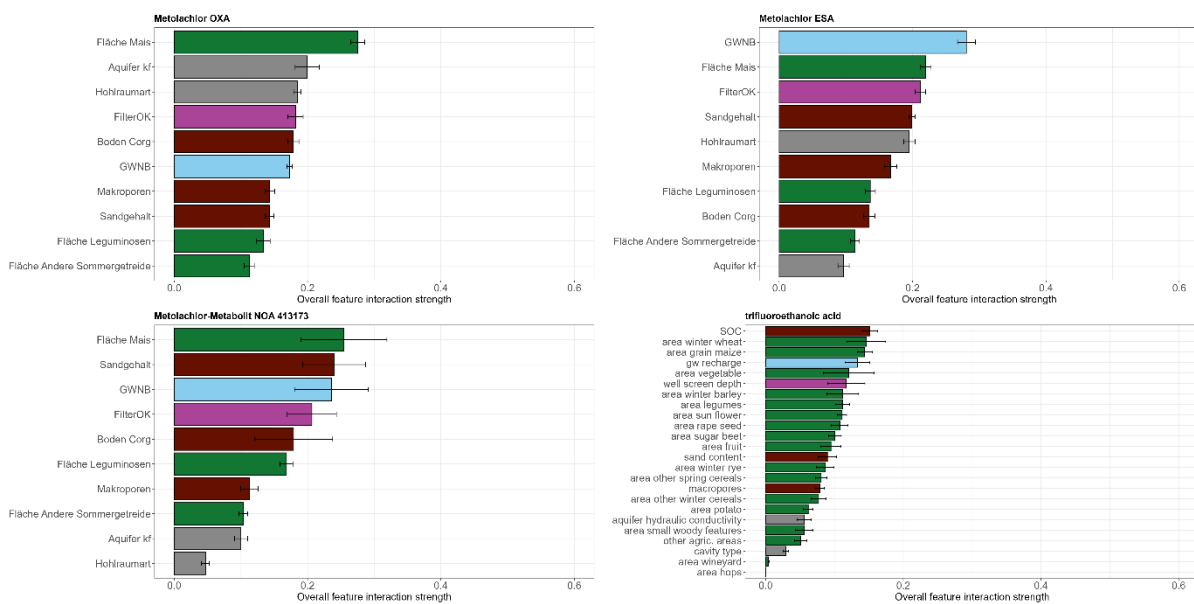
In Abbildung 73 und Abbildung 74 sind die übergreifenden Feature-Interaktionsstärken pro Stoff dargestellt. Die stärkste Interaktion von über 0,4 findet sich bei der Filteroberkante bei Desphenyl-Chloridazon, der Großteil der Interaktionen sind deutlich geringer. Bei TFA und den Metolachlor-Metaboliten übersteigt kein Feature eine Interaktionsstärke von 0,3. Insgesamt sind die Interaktionsstärken der Features bei den meisten Stoffen als gering bis moderat einzustufen, somit werden eher keine komplexen nichtlinearen Zusammenhänge gelernt.

Abbildung 73: Stärke der Feature-Interaktionen für vier ausgewählte Metabolite.



Quelle: eigene Darstellung BGR.

Abbildung 74: Stärke der Feature-Interaktionen für vier ausgewählte Metabolite.



Quelle: eigene Darstellung BGR.

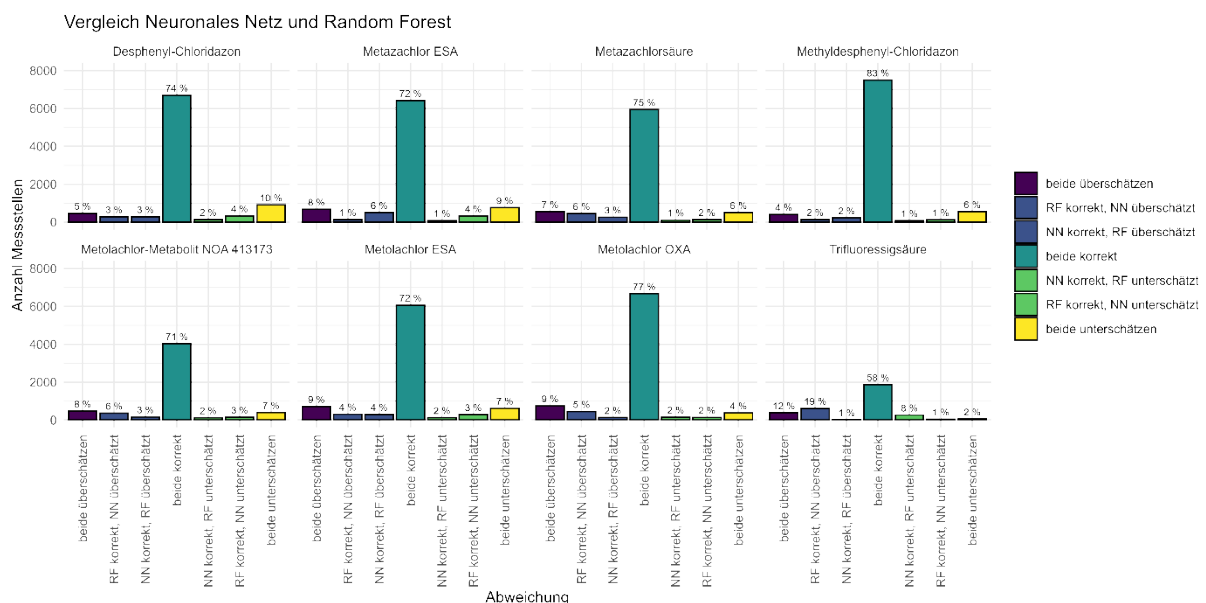
3.2.1.8 Vergleich der Vorhersage zwischen Random-Forest und dem Neuronalen Netz

Zusätzlich zur Analyse der Residuen in 3.2.1.6 sollen die Residuen im Folgenden zwischen den beiden Modellarchitekturen verglichen werden. Dabei ist von Interesse, ob abweichende Messstellen von beiden Architekturen systematisch über- bzw. unterschätzt werden oder ob die Architekturen an einzelnen Messstellen zu unterschiedlichen Vorhersagen führen. Die Abweichung zwischen tatsächlicher und vorhergesagter Konzentration wurde zwischen den RF- und den NN-Modellen an jeder Messstelle verglichen.

Bei den folgenden Ergebnissen handelt es sich jeweils um die binären Modelle der reduzierten Feature-Auswahl für die acht Metaboliten. Bei allen Metaboliten wird eine Mehrheit der Messstellen (58-83%) von beiden Algorithmen korrekt vorhergesagt. Je nach Stoff wird ein

gewisser Anteil an Messstellen von beiden überschätzt (4–12%) bzw. unterschätzt (0–10%). Weiterhin gibt es vier Kombinationen, bei denen jeweils eines der beiden Modelle korrekte Vorhersagen erzeugt, während das andere Modell über- bzw. unterschätzt. Wenn es zu dem Fall kommt, dass ein Modell korrekte, das andere inkorrekte Vorhersagen trifft, dann handelt es sich bei allen Stoffen häufiger um eine Überschätzung als um eine Unterschätzung. Diese vier Fälle liegen in der Summe je nach Stoff zwischen 6 und 13%, erreichen bei TFA jedoch 29%. Methyl-desphenyl-Chloridazon zeigt anteilig die geringste Abweichung zwischen RF- und NN-Modellen, während bei TFA die stärkste Abweichung zu verzeichnen ist. Hierbei ist der Fall, dass das RF-Modell korrekt vorhersagt, während das NN-Modell überschätzt, mit 19% relativ häufig.

Abbildung 75: Vergleich der Performance der zwei Algorithmen (Neuronale Netze und Random-Forest) anhand der Abweichung der tatsächlichen und vorhergesagten Konzentrationsklassen für die binären Modelle der reduzierten Feature-Auswahl.



Quelle: eigene Darstellung BGR.

Es ergibt sich an jeder Messstelle einer von sieben möglichen Fällen: Beide Algorithmen erzeugen korrekte Vorhersagen, beide über- bzw. unterschätzen oder eines der beiden Modelle erzeugt korrekte Vorhersagen, während das jeweils andere Modell über- oder unterschätzt.

3.2.2 Erweiterte Feature-Auswahl

Im Folgenden werden die Ergebnisse der Random-Forest-Modelle der erweiterten Feature-Auswahl der acht Metaboliten vorgestellt (2.2.3.6). Die verwendeten Features sind:

- ▶ Flächen pro Kulturklasse, für die der jeweilige Wirkstoff zugelassen ist
- ▶ Tiefe der Filteroberkante unter Geländeoberkante
- ▶ Sandgehalt im Boden
- ▶ Organischer Kohlenstoffanteil im Boden
- ▶ Makroporenklasse
- ▶ Langjährige Rate der Grundwasserneubildung
- ▶ Hohlraumart des Aquifers

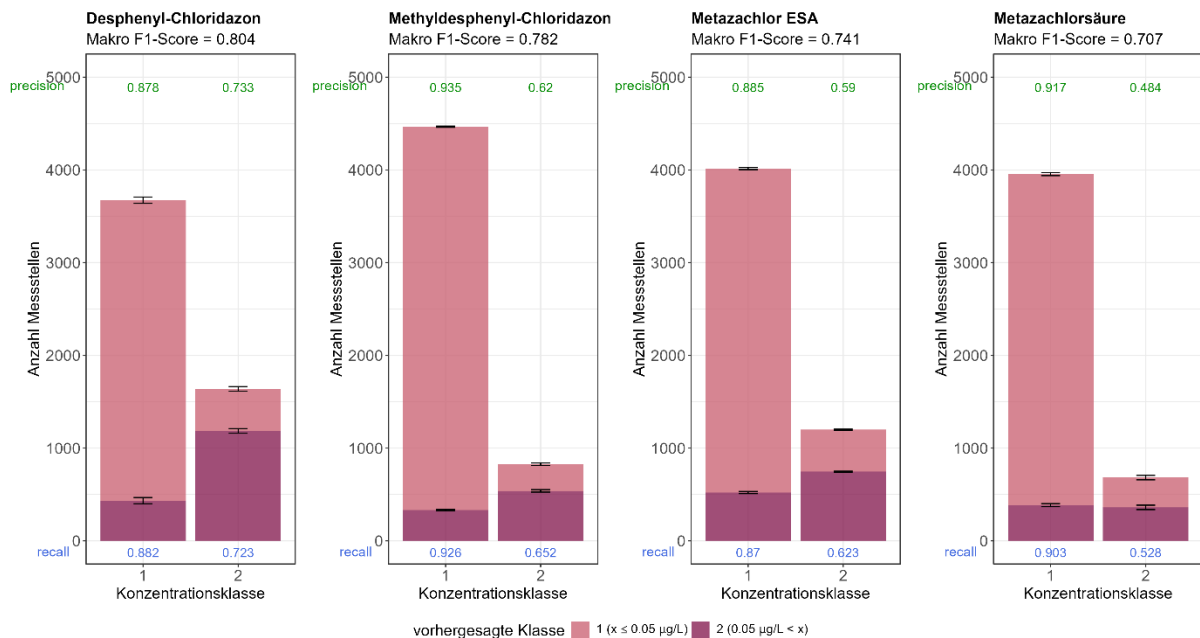
- ▶ Hydraulische Leitfähigkeitsklasse des Aquifers
- ▶ Hydrogeologischer Großraum
- ▶ langjähriges Mittel der Lufttemperatur
- ▶ mittlerer Sommerniederschlag
- ▶ mittlerer Winterniederschlag
- ▶ Verhältnis aus Niederschlag April-September/Okttober-März
- ▶ Geländehöhe
- ▶ Hangneigung
- ▶ Gesteinsart
- ▶ Geochemischer Typ
- ▶ Aus MOHP, abstrahierte Abstände zu Einzugsgebietsgrenzen und Gewässer für Strahler-
Ordnung 1 und 2 (DSD_1, DSD_2, LP_1, LP_2, SD_1, SD_2)
- ▶ Schluffgehalt im Boden
- ▶ Tongehalt im Boden
- ▶ Ackerfläche im Puffer (kulturunspezifisch)
- ▶ Sickerwasserrate
- ▶ mittlere Nitratkonzentration an der Messstelle
- ▶ mittlere Ammoniumkonzentration
- ▶ mittlere Magnesiumkonzentration
- ▶ mittlere Sulfatkonzentration
- ▶ mittlere Chloridkonzentration
- ▶ mittlere Kaliumkonzentration
- ▶ mittlere Natriumkonzentration

3.2.2.1 Vorhersagegüte

Die Modellgüte der Random-Forest Modelle für die erweiterte Feature-Auswahl wird in Abbildung 76 und Abbildung 77 gezeigt. Die Anzahl der Messstellen pro Klasse ist deutlich geringer als in den Modellen für die reduzierte Feature-Auswahl, da nur Messstellen mit vollständigen Daten für eine größere Anzahl an Features in Frage kommen. Das Verhältnis zwischen der Anzahl der Messstellen in den zwei Konzentrationsklassen ist unverändert zugunsten von Konzentrationsklasse 1. Auch zeigt sich, wie zu erwarten ist, dass Precision und Recall der selteneren Klasse geringer sind. Der Makro-F1-Score hat sich jedoch bei allen Stoffen gegenüber den Modellen mit der reduzierten Feature-Auswahl verbessert und liegt durchgehend über 0,7. Der Makro-F1-Score von Desphenyl-Chloridazon liegt über 0,8.

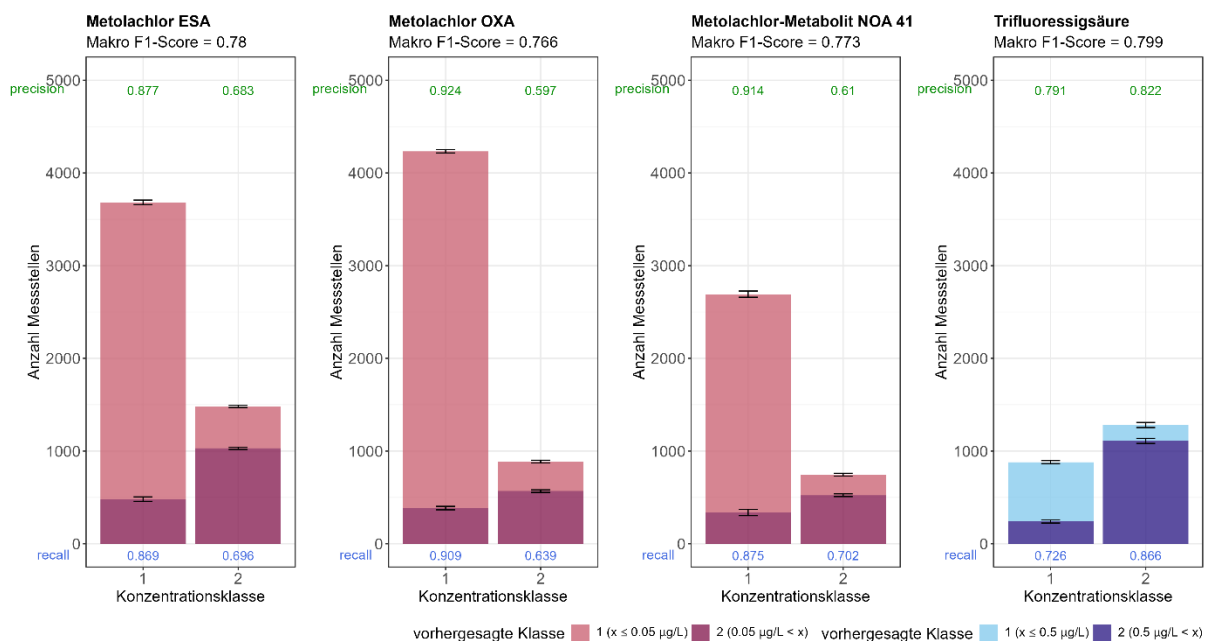
Der Vergleich der Modellgüten der zwei Varianten der Feature-Auswahl für das Neuronale Netz wird in Tabelle 5 gezeigt. Sowohl die RF- als auch die NN-Modelle erzielen höhere Macro-F1-Scores in den Modellen der erweiterten Feature-Auswahl im Vergleich zur reduzierten Feature-Auswahl.

Abbildung 76: Modellgüte der Random-Forest Modelle für die erweiterte Feature-Auswahl für zwei Chloridazon- und zwei Metazachlor-Metaboliten.



Quelle: eigene Darstellung BGR.

Abbildung 77: Modellgüte der Random-Forest Modelle für die erweiterte Feature-Auswahl für drei Metolachlor-Metaboliten und die Trifluoressigsäure.



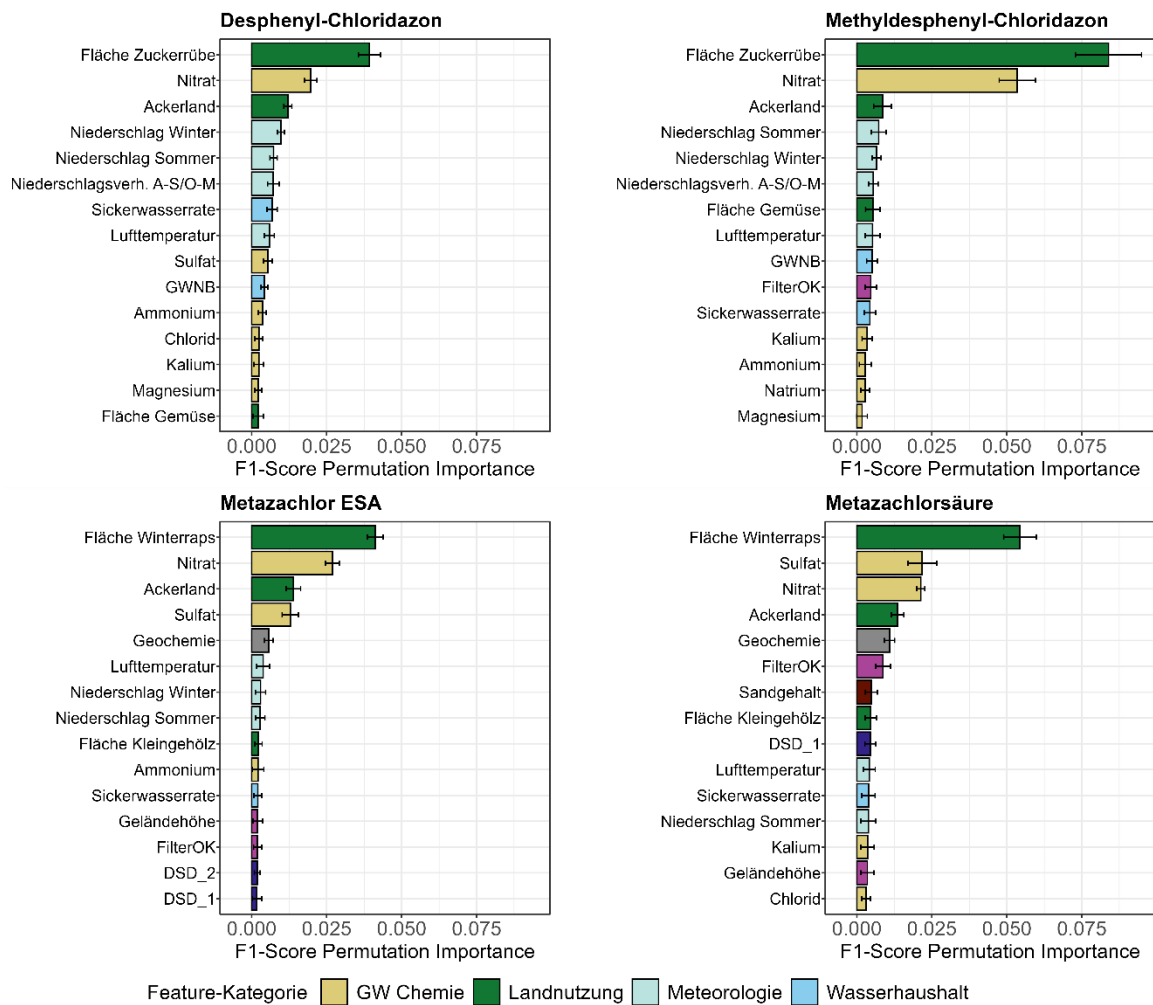
Quelle: eigene Darstellung BGR.

3.2.2.2 Featuregewichtung (Feature Importance)

In Abbildung 78 und Abbildung 79 werden für alle acht Metabolitenmodelle die jeweils 15 wichtigsten Features nach ihrer Gewichtung gemäß der Permutation Importance geordnet dargestellt. Das wichtigste Feature ist bei den meisten Metaboliten, wie auch bei der reduzierten Feature-Auswahl, die jeweilige Hauptkultur. Bei Metolachlor ESA teilen sich Kalium und die Maisfläche den ersten Rang, da sich die Unsicherheit der Featuregewichtung überschneidet. Bei TFA finden sich der mittlere Nitratgehalt, die Tiefe der Filteroberkante und der mittlere Ammoniumgehalt an vorderster Stelle. Bei den beiden Chloridazon-Metaboliten und Metazachlor ESA findet sich Nitrat deutlich an zweiter Stelle, bei der Metazachlorsäure sind der Sulfat- und der Nitratgehalt gleichauf an zweiter Position. Hinsichtlich der anorganischen Grundwasserchemie finden sich unterschiedliche Gewichtungen zwischen den Stoffen: Bei den Chloridazon-Metaboliten findet sich nur Nitrat in den oberen Rängen, weitere Parameter spielen eine zu vernachlässigende Rolle. Bei den Metazachlor-Metaboliten gesellt sich zu Nitrat auch der Sulfatgehalt, während dies bei den Metolachlor-Metaboliten für den Kaliumgehalt der Fall ist. Nach diesen Features ist bei den Chloridazon- und den Metazachlor-Metaboliten noch das Feature Ackerland zu erwähnen. Features von mittlerer Importance sind bei den Chloridazon-Metaboliten Klimafeatures, wie der mittlere Sommer- bzw. Winterniederschlag oder die Lufttemperatur. Letztere erreicht bei Metazachlor ESA und bei den drei Metolachlor-Metaboliten einen mittleren Rang. Bei TFA ist der Sommerniederschlag und die Anbaufläche an Ackerland zu erwähnen. Weitere nennenswerte Features sind die Geländehöhe bei Metolachlor ESA und OXA, der hydrogeologische Großraum bei Metolachlor ESA und OXA sowie die Geochemie des Aquifers bei den Metazachlor-Metaboliten.

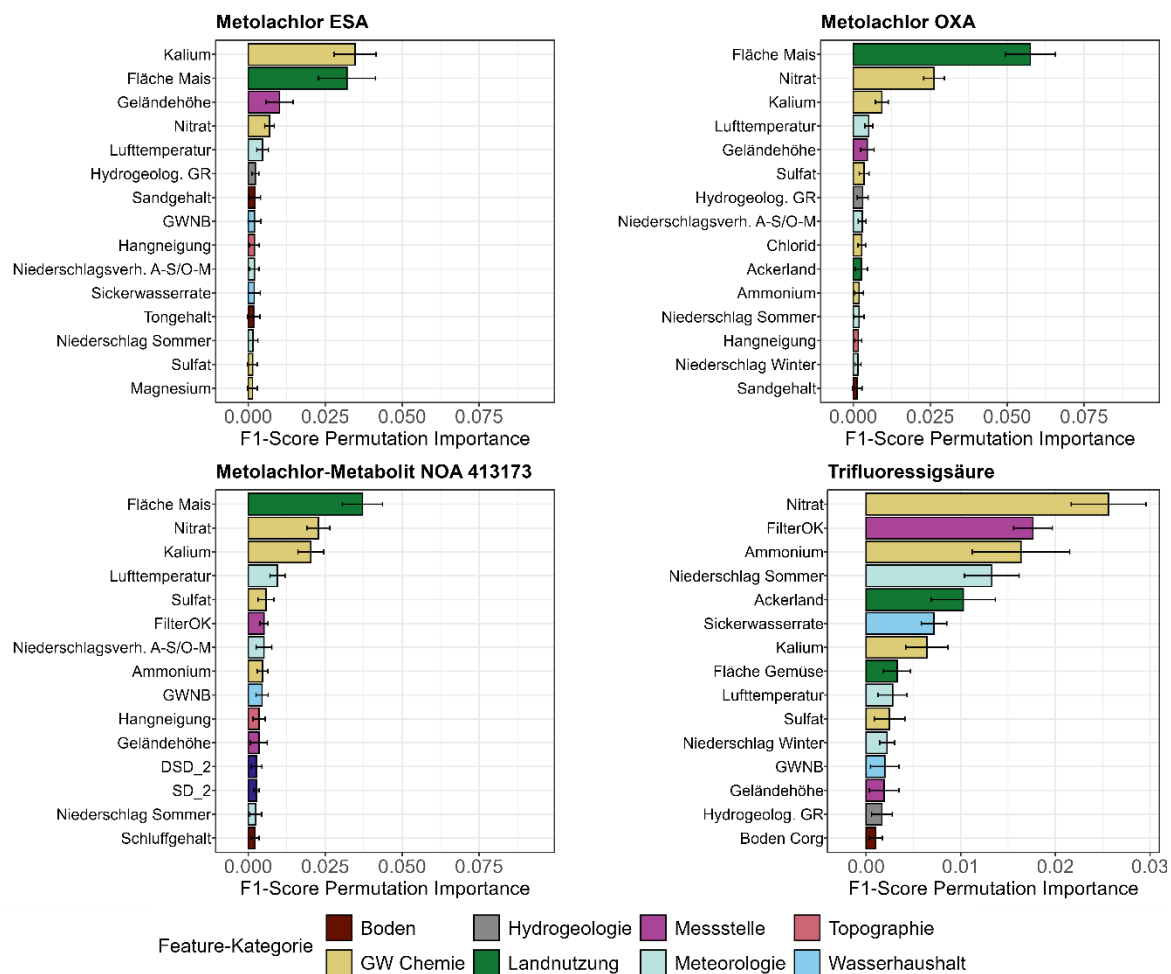
Einige Feature-Kategorien sind von deutlich geringerer Gewichtung oder finden sich nicht in den Top15 wieder. Bei den meisten Stoffen sind die Bodenfeatures relativ unbedeutend, alleinig bei Metazachlorsäure erreicht der Sandgehalt eine Importance von größer Null, die jedoch von weiteren Features nicht zu unterscheiden ist. Makroporen, die hydraulische Leitfähigkeit und die Hohlraumart des Aquifers tauchen nicht mehr auf. Die Grundwasserneubildung findet sich bei einem Teil der Stoffe mit geringer Importance noch in den Top15. Auch die Filteroberkante verliert bei den meisten Stoffen im Vergleich zu den Modellen mit reduzierter Feature-Importance deutlich an Bedeutung, bei Methyldephenyl-Chloridazon, Metazachlorsäure und TFA taucht die Filteroberkante in der Top15 auf. Nebenkulturen verlieren bei allen Stoffen deutlich an Relevanz.

Abbildung 78: Die jeweils 15 wichtigsten Features nach der Permutation Feature Importance für das binäre Random-Forest Modell der erweiterten Feature-Auswahl für Desphenyl-Chloridazon, Methyldesphenyl-Chloridazon, Metazachlor ESA und Metazachlorsäure.



Quelle: eigene Darstellung BGR.

Abbildung 79: Fortführung von Abbildung 78 für Metolachlor ESA, Metolachlor OXA, Metolachlor NOA 413173 und die Trifluoressigsäure.

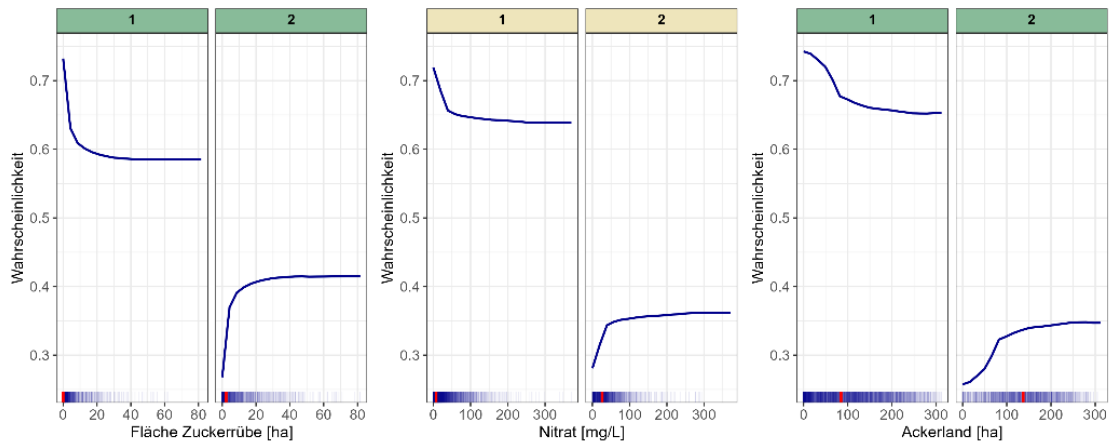


Quelle: eigene Darstellung BGR.

3.2.2.3 Feature-Effekte

Die Feature-Effekte der jeweils drei wichtigsten Features (gemäß Permutation-Importance) der acht Metaboliten sind ab Abbildung 80 dargestellt. Die Features, die bereits in den Modellen der reduzierten Feature-Auswahl vorkommen, zeigen grundsätzlich das gleiche Verhalten wie zuvor. Das Augenmerk soll in dieser Analyse daher auf die zusätzlichen Features gelegt werden. Nitrat, Kalium und Sulfat verhalten sich hinsichtlich ihrer Effektkurven wie die Landnutzungsfeatures (Hauptkultur). Mit zunehmenden mittleren Konzentrationen im Grundwasser steigt die Wahrscheinlichkeit für Konzentrationsklasse 2. Die Stärke des Effekts ist bei Nitrat vergleichbar mit der der Hauptkultur bei den meisten Metaboliten. Meist ist der Anstieg der Effektkurve anfangs relativ steil, während er bei Metolachlor ESA flacher ausfällt. Kalium und Sulfat zeigen weniger ausgeprägte Effekte als Nitrat. Das Feature Ackerland zeigt einen graduelleren Anstieg als die Hauptkultur oder als Nitrat und in der Regel einen schwächeren Effekt. Die Wertespanne, über die sich der Effekt erstreckt, ist bei beiden Features relativ klein. Bei Metolachlor OXA fällt die Wahrscheinlichkeit für Konzentrationsklasse 2 mit steigender Geländehöhe. Auch im Fall von TFA (Abbildung 87) zeigt Nitrat den bereits beschriebenen Effekt. Die Effektkurve Filteroberkante zeigt ebenfalls den gleichen Zusammenhang wie bei allen Stoffen bisher. Der mittlere Ammoniumgehalt zeigt für kleine Konzentrationswerte einen zu Nitrat gegensätzlichen Trend, der vorwiegende Teil der Kurve verläuft jedoch flach.

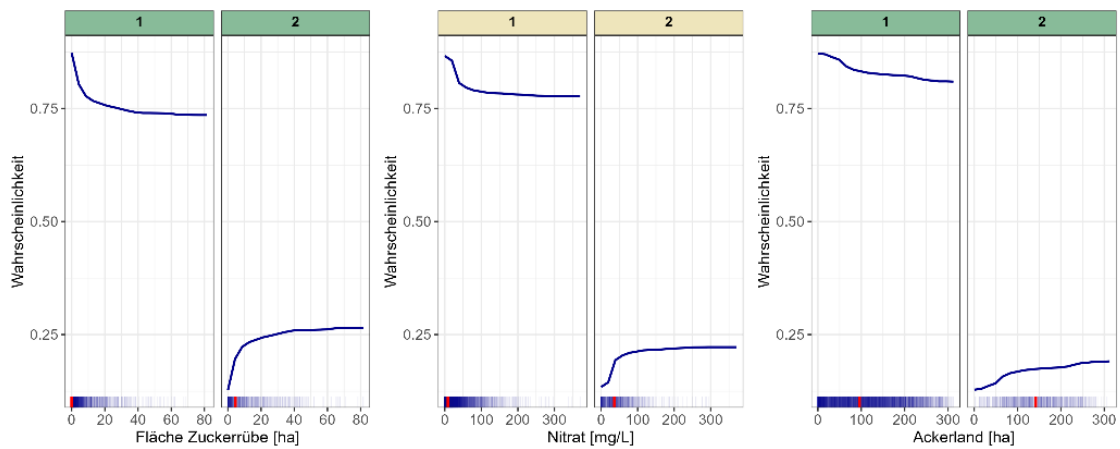
Abbildung 80: Feature-Effekte von Desphenyl-Chloridazon im binären Random-Forest Modell der erweiterten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.



Quelle: eigene Darstellung BGR.

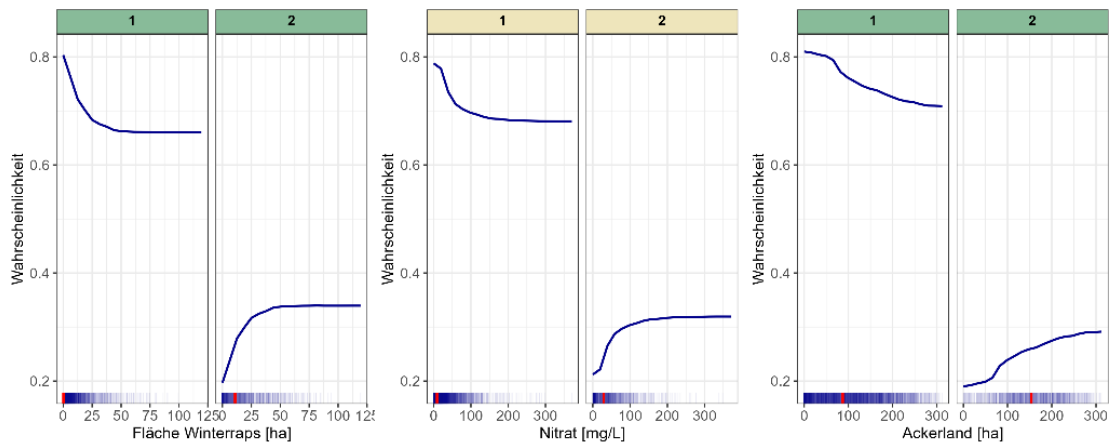
Die Y-Achse stellt die Wahrscheinlichkeit der Vorhersage von jeweils Konzentrationsklasse 1 und 2 dar, die X-Achse zeigt den Wertebereich des jeweiligen Features an. Die blaue Linie stellt die gemittelte Effektkurve dar, der hellblaue Schatten verdeutlicht die Variabilität anhand der unterschiedlichen Seeds. Die Effekte der einzelnen Ausprägungen der kategorischen Features sind als Balkenplots dargestellt, die Fehlerbalken veranschaulichen die Seeds.

Abbildung 81: Feature-Effekte von Methyl-desphenyl-Chloridazon, Fortführung vorheriger Abbildungen.



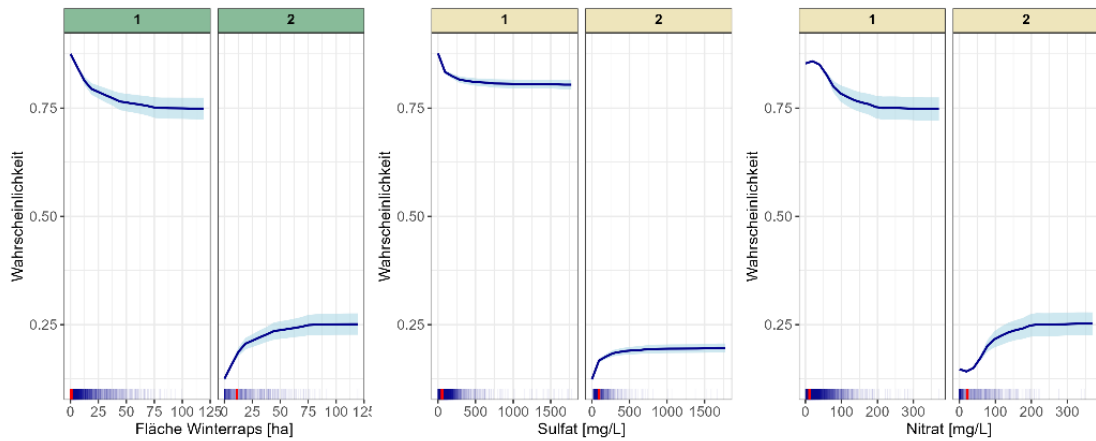
Quelle: eigene Darstellung BGR.

Abbildung 82: Feature-Effekte von Metazachlor ESA, Fortführung vorheriger Abbildungen.



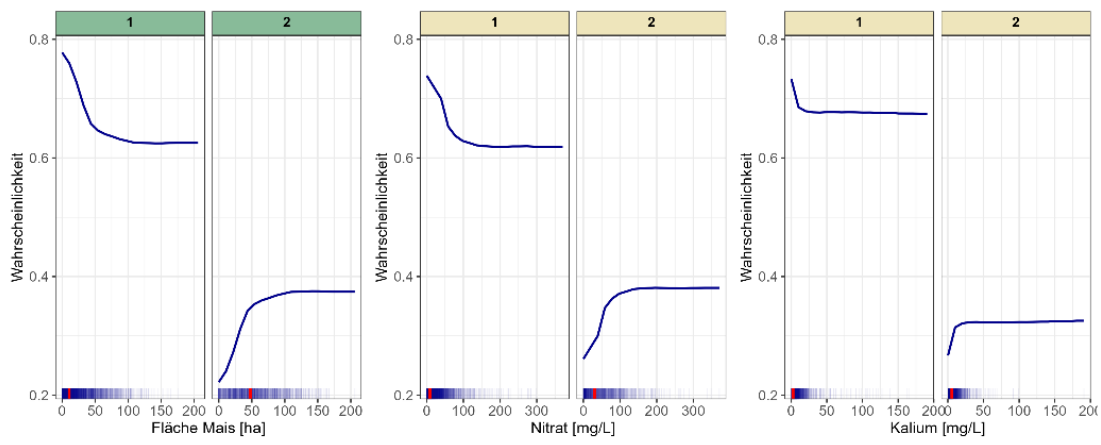
Quelle: eigene Darstellung BGR.

Abbildung 83: Feature-Effekte der Metazachlorsäure, Fortführung vorheriger Abbildungen.



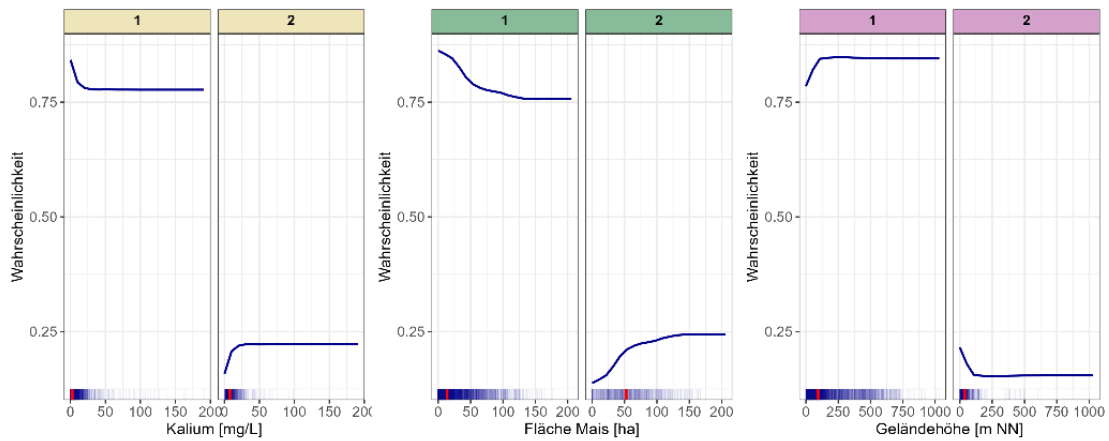
Quelle: eigene Darstellung BGR.

Abbildung 84: Feature-Effekte von Metolachlor ESA, Fortführung vorheriger Abbildungen.



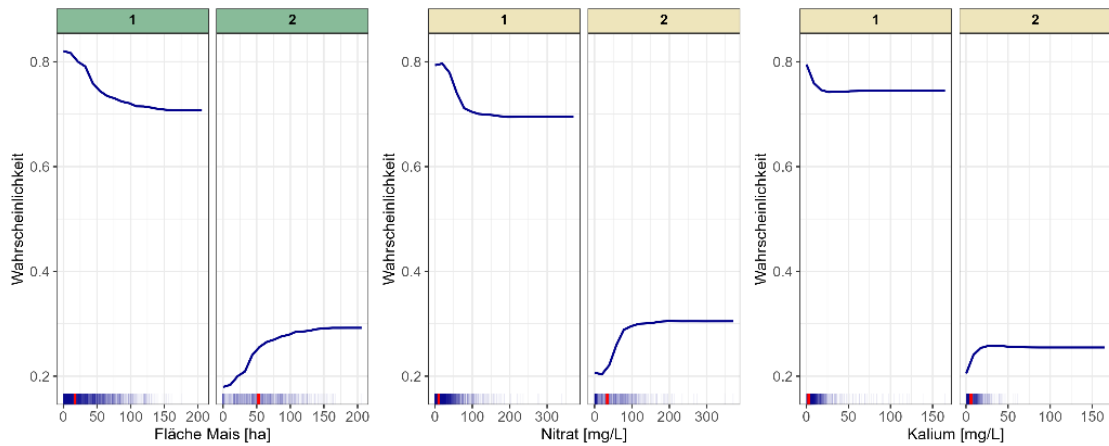
Quelle: eigene Darstellung BGR.

Abbildung 85: Feature-Effekte von Metolachlor OXA, Fortführung vorheriger Abbildungen.



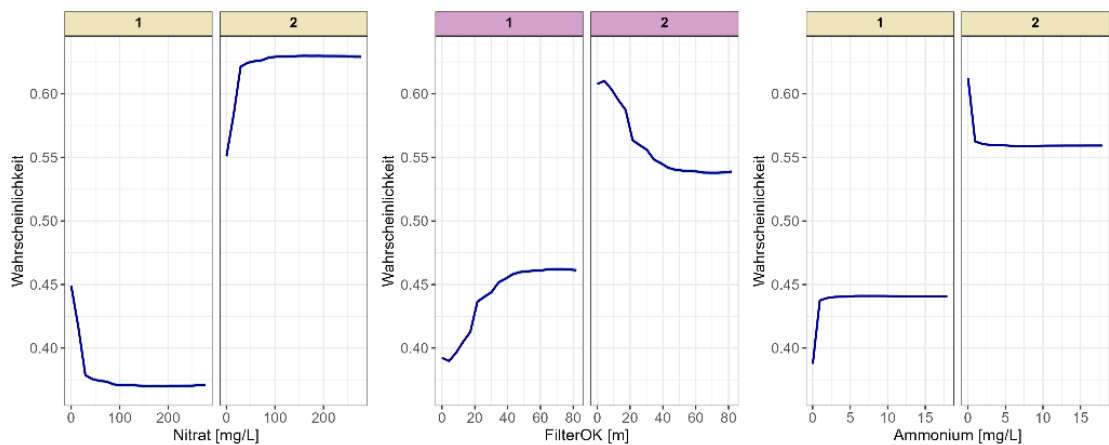
Quelle: eigene Darstellung BGR.

Abbildung 86: Feature-Effekte von NOA 413173, Fortführung vorheriger Abbildungen.



Quelle: eigene Darstellung BGR.

Abbildung 87: Feature-Effekte von TFA, Fortführung vorheriger Abbildungen.



Quelle: eigene Darstellung BGR.

3.2.3 Zusammenfassung

Es wurden binäre Klassifikationsmodelle mittels zweier Architekturen für die Metaboliten Desphenyl-Chloridazon, Methyl-desphenyl-Chloridazon, Metolachlor ESA, Metolachlor OXA und Metolachlor NOA 413173 sowie Metazachlor ESA, Metazachlorsäure und Trifluoressigsäure (TFA) trainiert. Als Features (erklärende Variablen) wurden die Fläche der Haupt- bzw. Nebenkulturen im 1-km Radius um die Messstelle, die Tiefe der Filter-Oberkante, die mittlere Grundwasserneubildung, der Gehalt an Sand bzw. organischem Kohlenstoff im Boden, die Makroporenklasse sowie die hydraulische Leitfähigkeit bzw. die Hohlraumart des Aquifers verwendet. Mit der binären Klasseneinteilung (1) $\leq 0,05 \mu\text{g L}^{-1}$; 2) $> 0,05 \mu\text{g L}^{-1}$) konnten für die meisten Stoffe ein Macro-F1-Score von ca. 0,7 (0.674-0.77) und somit gute Modellgüten erreicht werden. Bei TFA erwies sich die Klasseneinteilung (2. Klasse $> 0,5 \mu\text{g L}^{-1}$) als geeigneter. Insgesamt wird der nachteilige Einfluss der Ungleichverteilung der Klassengrößen auf die Vorhersagegüte sichtbar. Die Modellgüte ist bei den meisten Stoffen auch bei der selteneren zweiten Klasse ausreichend, bei Metazachlor ESA sind die Gütemaße der zweiten Klasse grenzwertig. Unabhängig vom Algorithmus oder der Feature-Importance-Methode wurde bei fast allen Metaboliten die jeweilige Hauptkultur als das mit Abstand wichtigste Feature identifiziert. Als weitere, nachfolgende Features sind die Filter-Oberkante, der Sandgehalt, die Grundwasserneubildung und z.T. auch Nebenkulturen zu nennen. Die Reihenfolge dieser Features kann sich nach Methode etwas unterscheiden. Bei TFA wird die Filter-Oberkante und der Sandgehalt als am wichtigsten eingestuft, es folgen zahlreiche Kulturen. Den hydrogeologischen Features und der Makroporenklasse wurde für alle Stoffe eine geringe Feature-Gewichtung zugeordnet. Bei allen Stoffen findet sich der Zusammenhang, dass eine größere Fläche der Hauptkultur mit der höheren Konzentrationsklasse und eine tiefere Filter-Oberkante mit der niedrigeren Konzentrationsklasse assoziiert ist. Weitere Features wie der Sandgehalt zeigen Unterschiede zwischen den Stoffen. In Kapitel 4 wird darauf eingegangen, ob diese Unterschiede mit der Verteilung dieser Features in den jeweiligen Anbaubereichen in Zusammenhang gebracht werden können. Messstellen, an denen die Modelle zu Überschätzungen der Konzentrationen kommen, weisen z.T. geringere Filtertiefen und höhere Flächenanteile der jeweiligen Hauptkultur auf. Bei Messstellen, an denen es zu Unterschätzungen kommt, ist der gegensätzliche Zusammenhang sichtbar. Die räumliche Verteilung der Messstellen und ihrer Vorhersagen zeigt, dass bundesweit Messstellen korrekt vorhergesagt werden, Unter- und Überschätzungen finden sich auch in jedem Bundesland. Die Architekturen NN und RF stimmen in ihrer Vorhersage an den meisten Messstellen überein, dies betrifft korrekte Vorhersagen als auch Über- oder Unterschätzungen. Der Anteil an Messstellen, an denen eine Architektur unter- und die andere überschätzt, ist relativ gering. Aufgrund von geringer bis moderater Interaktionsstärken der Features bei den meisten Stoffen ist davon auszugehen, dass die Modelle eher additive als nichtlineare Zusammenhänge zwischen den Features lernen. Die Implikationen aus diesen Ergebnissen werden in Kapitel 4 behandelt.

Die Hinzunahme weiterer Features führte bei allen Stoffen zu einer Reduzierung der verfügbaren Messstellenanzahl. Die Modellgüten waren dennoch höher als bei der reduzierten Feature-Auswahl. Die jeweilige Hauptkultur ist bei fast allen Stoffen nach wie vor das wichtigste Feature. Ebenfalls sehr relevant sind die mittleren Gehalte an Nitrat, Sulfat und Kalium im Grundwasser. Diese Features weisen einen mit der Anbaukultur vergleichbaren Einfluss auf die Vorhersage der Konzentrationsklassen auf.

3.2.4 Wirkstoffmodelle für Stoffgruppen und Einzelstoffe

Die zuvor behandelten Metaboliten-Modelle repräsentieren relativ mobile Abbauprodukte, die im Boden entstehen. Um auch Wirkstoffe beleuchten zu können, die oberirdisch appliziert

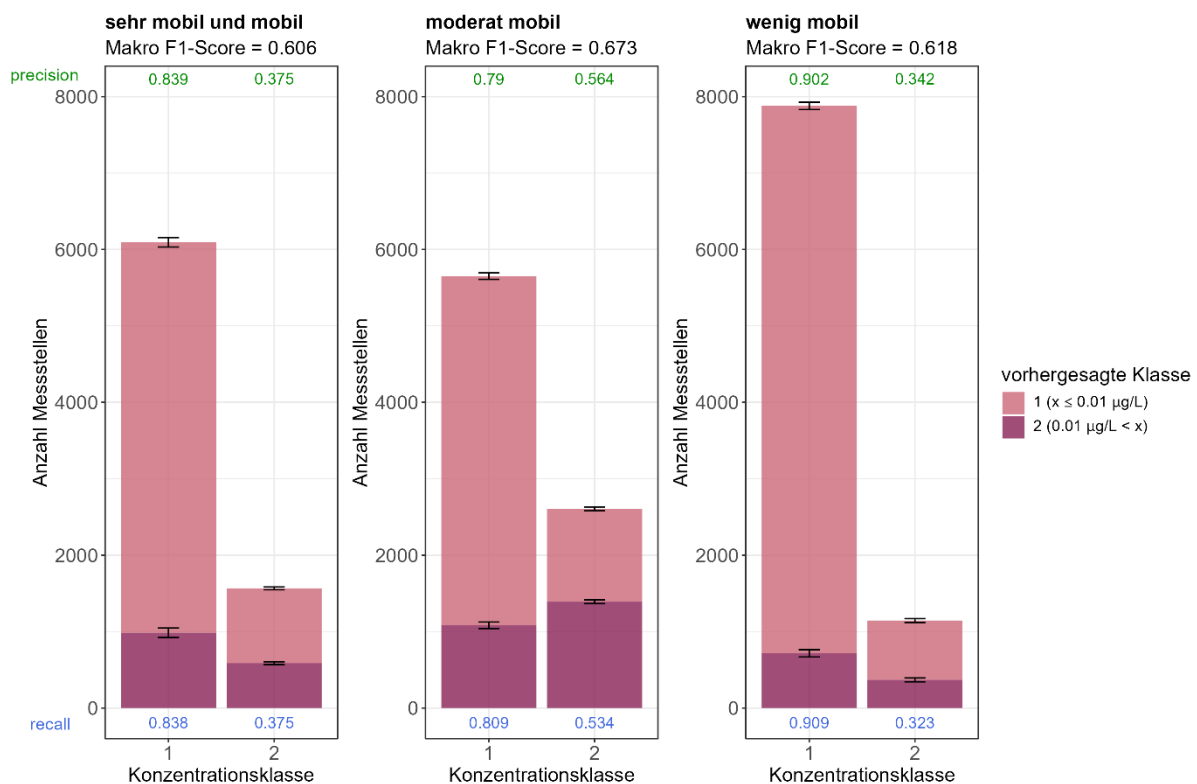
werden und andere Eigenschaften aufweisen können, wurden Einzelstoffmodelle vereinzelt gerechnet. Allerdings war dies letztlich aufgrund des hohen Zensuranteils für die Mehrheit der Wirkstoffe nicht möglich. Daher wurde eine Möglichkeit gesucht, Wirkstoffe in Gruppen zusammenzufassen und Modelle für diese Stoffgruppen zu trainieren. Wirkstoffgruppen wurden anhand ihrer Mobilität (K_{Foc}) in 3 Kategorien eingeordnet (sehr mobil und mobil, moderat mobil und wenig mobil) und die Klassengrenze $0,01 \mu\text{g L}^{-1}$ in einer binären Klassifikation angewandt (siehe 2.2.3.6). Einzelstoffmodelle wurden für Bentazon, Atrazin und Simazin generiert. Wirkstoff- bzw. Wirkstoffgruppenmodelle wurden ausschließlich mit Random-Forest-Modellen erzeugt. Im Folgenden werden die Ergebnisse der reduzierten Feature-Auswahl gezeigt.

3.2.4.1 Reduzierte Feature-Auswahl

3.2.4.1.1 Vorhersagegüte

Das Wirkstoffgruppen-Modell für sehr mobile und mobile Wirkstoffe erreicht einen F1-Score von 0,606, das Modell für moderat mobile Wirkstoffe erreicht 0,673 und das Modell für wenig mobile Wirkstoffe 0,618 (Abbildung 88). Precision und Recall der zweiten Klasse (Überschreitungen $> 0,01 \mu\text{g L}^{-1}$) liegen bei den sehr mobilen und den wenig mobilen Wirkstoffen jeweils unter 0,4 und somit ist die Vorhersagegüte schlechter als Zufall. Moderat mobile Wirkstoffe zeigen eher ausgeglichene Gütemaße in den beiden Konzentrationsklassen.

Abbildung 88: Die Modellgüte der binären Random-Forest-Modelle der reduzierten Feature-Auswahl für drei anhand der Mobilität eingeteilte Wirkstoffgruppen.



Quelle: eigene Darstellung BGR.

Die Abbildungen der genannten Einzelstoffmodelle werden hier nicht gezeigt. Das Modell für Bentazon (mobil) erzielt einen Macro-F1-Score von 0,655, allerdings liegt die Precision bei 0,445 und der Recall bei 0,44 hinsichtlich der Konzentrationsklasse 2. Das Modell für Atrazin (moderat mobil) erzielt einen Macro-F1-Score von 0,708, mit eine Precision von 0,819 bzw. 0,597 sowie

einem Recall von 0,807 bzw. 0,613. Bei Simazin (wenig mobil) ergibt sich ein Macro-F1-Score von 0,639, basierend auf einer Precision von 0,902 bzw. 0,383 sowie einem Recall von 0,906 bzw. 0,372. Somit erreicht nur das Atrazin-Modell Metriken für Konzentrationsklasse 2, die besser sind als eine zufällige Zuordnung.

3.2.4.1.2 Feature Importance

Aufgrund der geringen Modellgüte (bezüglich der Konzentrationsklasse 2) sind die Ergebnisse zur Feature Importance insbesondere für die Modelle der sehr mobilen und mobilen bzw. wenig mobilen Wirkstoffe nicht ausreichend verlässlich. Daher ist bei der Interpretation der Feature-Importance Vorsicht geboten. Für die mobilen und sehr mobilen Wirkstoffe wurde die Grundwasserneubildung, Ackerland, die Filter-Oberkante gemäß Permutation am wichtigsten eingestuft, ohne nennenswerte Unterschiede zwischen diesen drei Features. Nachrangig werden die Makroporenklasse, der Sandgehalt und Bodenkohlenstoffgehalt gelistet. Die Feature-Importance der Aquifer-kf-Klasse und der Hohlraumart wird als gering bewertet. Für die wenig mobilen Wirkstoffe wurde der Sandgehalt am wichtigsten eingestuft, danach folgen die Grundwasserneubildung, Bodenkohlenstoff und die Filter-Oberkante. Die Unterschiede in der Feature-Importance zwischen diesen Features sind gering. Deutlich weniger wichtig wurden jedoch Ackerland, Makroporen, die Hohlraumart und die Aquifer-kf-Klasse eingestuft.

Aus genannten Gründen der Modellgüte wird im Folgenden nur das Modell für moderate mobile Wirkstoffe eingehender betrachtet (Abbildung 90). Die Feature Importance gemäß Permutation (links) ordnet dem Sandgehalt den höchsten Rang zu. Nachfolgende Features wie der organische Kohlenstoffgehalt im Boden, die hydraulische Durchlässigkeit des Aquifer und die Grundwasserneubildung weisen eine ähnliche Gewichtung auf. In absteigender Reihenfolge sind die Filteroberkante, die Hohlraumart, das Ackerland und die Makroporenklasse zu nennen.

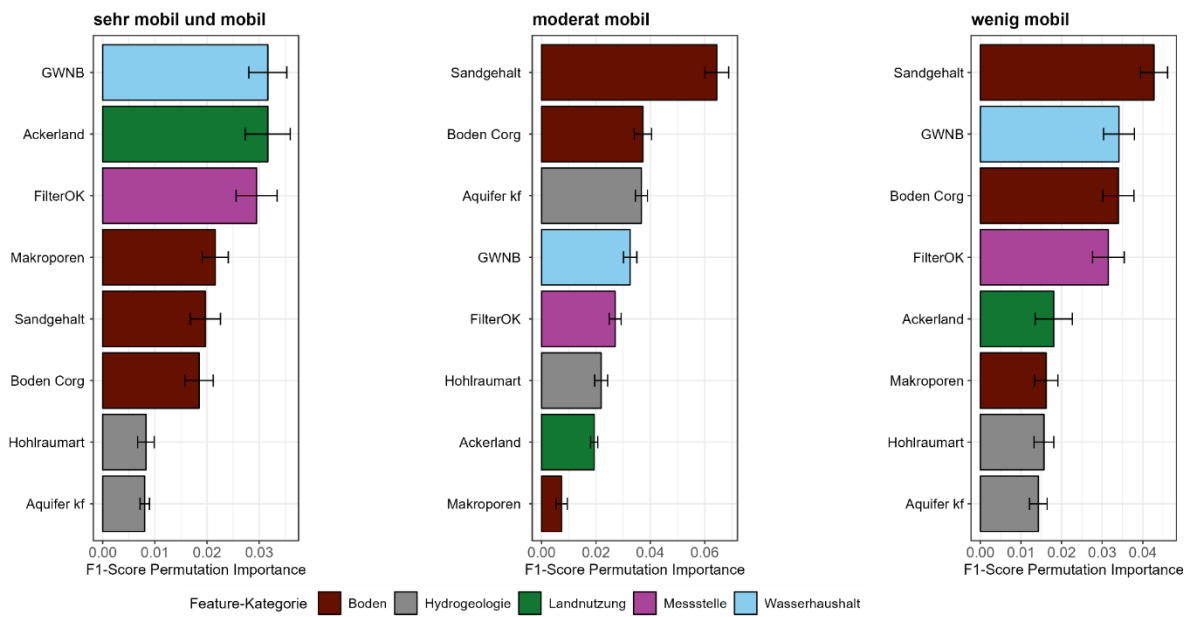
Die Feature Importance gemäß Gini impurity (Mitte) ordnet die Grundwasserneubildung als wichtigstes Feature ein, dicht gefolgt von der Anbaufläche an Ackerland und der Filteroberkante mit vergleichbarer Importance. Danach werden der Sandgehalt und der organische Kohlenstoff im Boden eingeordnet. Mit deutlich geringerer Importance folgen Aquifer-kf und die Hohlraumart.

Die korrigierte Gini impurity (rechts) zeigt den Sandgehalt an erster Stelle, Aquifer-kf an zweiter Stelle. Danach folgen die Filteroberkante, die Grundwasserneubildung, der organische Kohlenstoff und die Hohlraumart mit vergleichbarer, schwer unterscheidbarer Gewichtung. An vorletzter Stelle wird die Anbaufläche an Ackerland genannt. Alle drei Methoden weisen den Makroporen den letzten Rang zu. Die Rangordnung nach Permutation Importance und korrigierter Gini impurity zeigt deutliche Unterschiede zu den Metaboliten-Modellen: Boden-Features und Aquifer-kf ranken hier am höchsten, während der Anteil an Ackerland im Puffer als deutlich weniger bedeutend eingestuft wird. Die Rangordnung nach Gini impurity zeigt eine Abfolge, die der der Metabolitenmodelle stark ähnelt.

Unterschiede zwischen der Gini impurity und der Permutation Importance sind bei den Metabolitenmodellen deutlich geringer ausgeprägt. Eine mögliche Ursache für die Unterschiede zwischen den Methoden bei den Modellen für die Wirkstoffgruppen könnte eine aufgrund der zahlreichen Wirkstoffe weniger eindeutige Stellung der Features sein. Möglicherweise spielt daher der Bias der beiden Methoden hier verstärkt eine Rolle. Wenn wiederum alle Feature Importance Methoden ein ähnliches Ranking zeigen, kann man daraus schlussfolgern, dass die Importance für das jeweilige Modell eindeutig genug ist, so dass die Nachteile einzelner Methoden weniger ins Gewicht fallen (wie bei den Metabolitenmodellen der Fall). Die Gini impurity hat die Tendenz, numerische Features (mit vielen unterschiedlichen Werten) oder kategorische Features mit vielen möglichen Kategorien als wichtiger zu bewerten, während

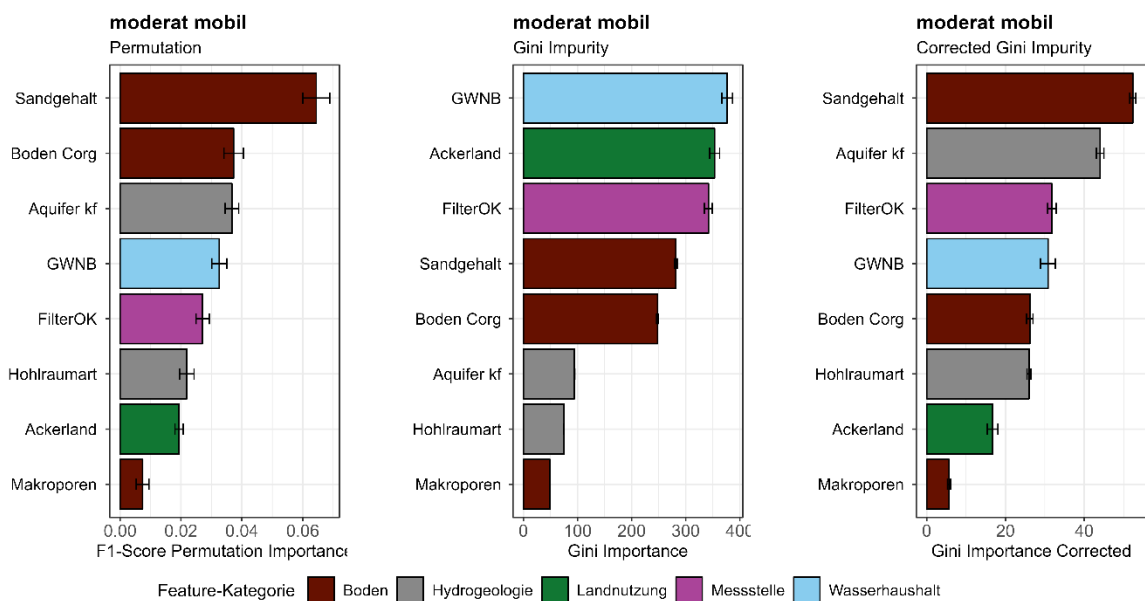
Features mit wenigen Kategorien (Makroporen (3); Aquifer-kf (3), Hohlraumart (4)) eine geringere Gewichtung zugeordnet wird. Dieser Bias kommt bei Permutation so nicht vor. Permutation hat jedoch den Nachteil, dass korrelierte Features eine höhere Bedeutung erhalten können. In diesem Fall ist die Rangordnung nach korrigierter Gini impurity der Permutation näher. Auch die Stärke der Feature-Effekte in den PDP-Plots verweist auf die Wichtigkeit der Features (nachfolgendes Teilkapitel). Da Permutation, korrigierte Gini impurity und die Feature-Effekte die Wichtigkeit der Features wie bspw. Aquifer-kf unterstreichen, können diese Features trotz der Unsicherheiten als relevant interpretiert werden.

Abbildung 89: Permutation-Importance für die drei Wirkstoffgruppenmodelle nach Mobilität (reduzierte Feature-Auswahl).



Quelle: eigene Darstellung BGR.

Abbildung 90: Permutation, Gini impurity und korrigierte Gini impurity exklusiv für die moderat mobilen Wirkstoffgruppenmodelle.



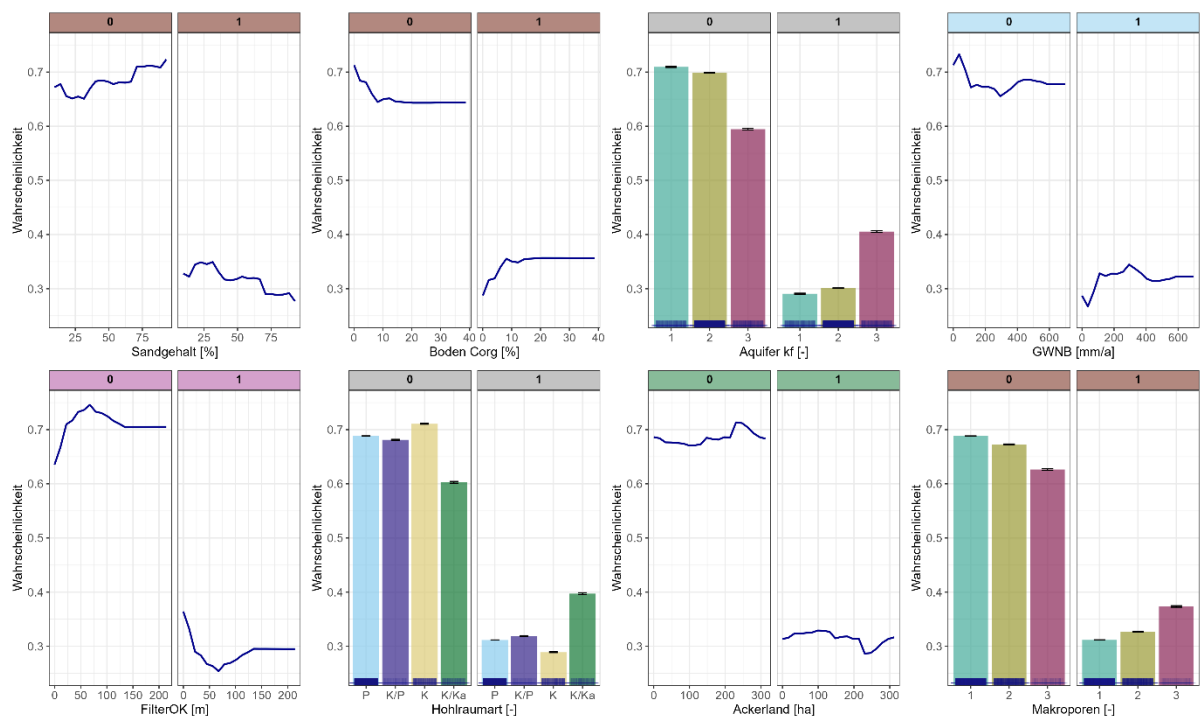
Quelle: eigene Darstellung BGR.

Bei Atrazin ist das wichtigste Feature gemäß Permutation die Grundwasserneubildung, gefolgt vom Sandgehalt und Aquifer-kf. Die Grundwasserneubildung und der Sandgehalt sind gemäß der Gini impurity und korrigierter Gini impurity relevant.

3.2.4.1.3 Feature-Effekte

Die Feature-Effekte der moderat mobilen Wirkstoffe (Abbildung 91) zeigen einen Abfall der Wahrscheinlichkeit der Vorhersage der Konzentrationsklasse 2 bei ansteigendem Sandgehalt. Beim organischen Kohlenstoffgehalt im Boden zeigt sich ein gegenläufiger Zusammenhang: ein Anstieg des C_{org} -Gehaltes geht mit einer Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 einher. Die Aquifer-kf-Klasse 3 ist mit einem Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 assoziiert. Der Effektverlauf der Grundwasserneubildung ist etwas erratisch, generell ist ein leichter Anstieg der Wahrscheinlichkeit für Konzentrationsklasse 2 mit ansteigender Grundwasserneubildung sichtbar. Die Filteroberkante zeigt einen Verlauf wie bei allen bisher gezeigten Modellen, ab ca. 70 m Tiefe der Filteroberkante scheint sich der Effekt zu einem gewissen Grad umzukehren. Die Hohlräumart Kluft/Karst ist mit einem Anstieg der Wahrscheinlichkeit von Konzentrationsklasse 2 assoziiert, genauso Makroporenklasse 3. Der Effekt der Anbaufläche an Ackerland ist als gering einzustufen. Die Feature-Effekte für Atrazin zeigen keine nennenswerten Unterschiede zu denen der moderat mobilen Wirkstoffgruppe.

Abbildung 91: Feature-Effekte der moderat mobilen Wirkstoffgruppe im binären Random-Forest Modell der reduzierten Feature-Auswahl. Partial-Dependence Plots wurden für jedes Feature und beide Konzentrationsklassen erzeugt.

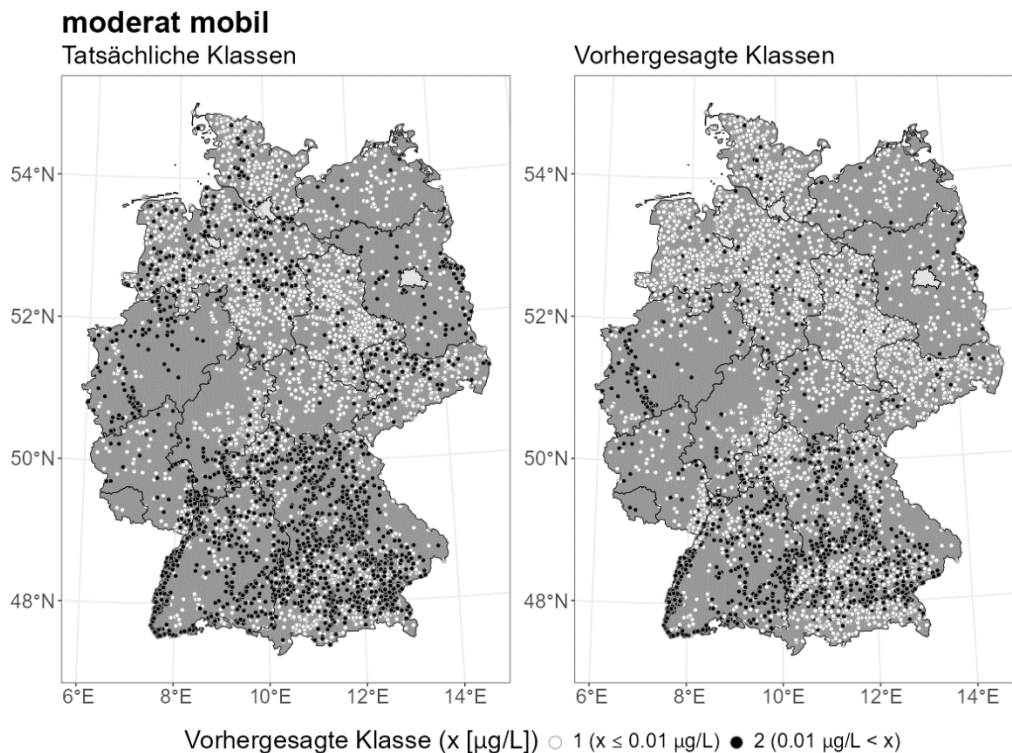


Quelle: eigene Darstellung BGR.

3.2.4.1.4 Räumliche Analysen

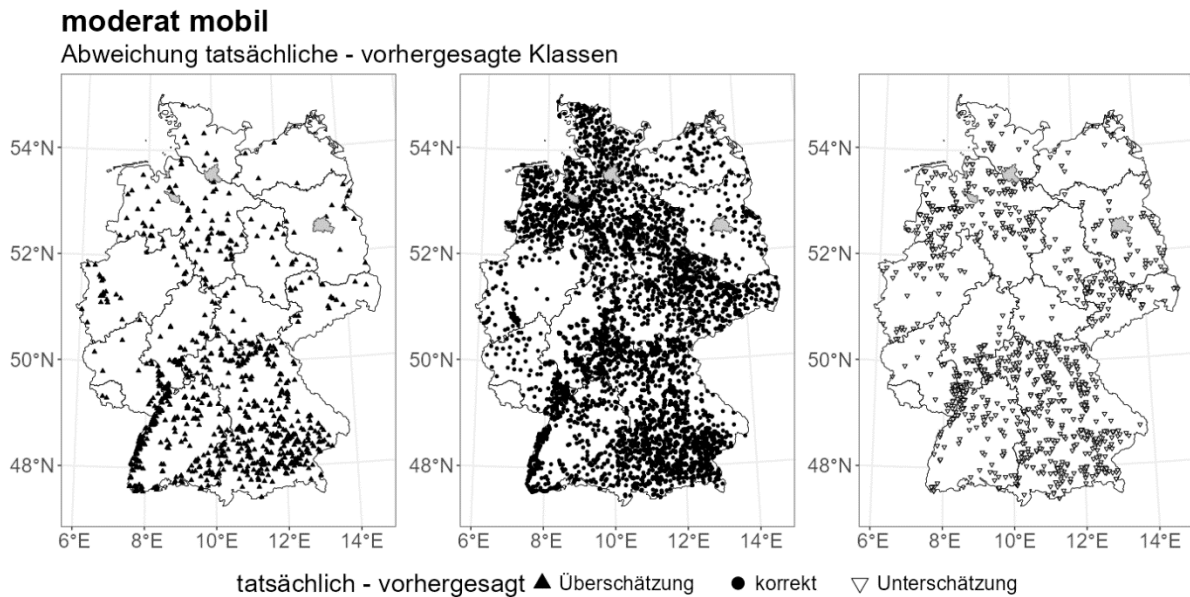
Aufgrund der besten Datengrundlage wird im Folgenden nur auf die Modellergebnisse der moderat mobilen Wirkstoffe der vorhergesagten Konzentrationsklassen eingegangen (Abbildung 92). In Süddeutschland erscheint es, dass absolut die Konzentrationsklasse 2 eher unter- bzw. überschätzt wird (Abbildung 93). Die Abweichung der tatsächlichen von der vorhergesagten Konzentrationsklasse der Grundwassermessstelle ist besonders in Norddeutschland als gut anzusehen. Für Schleswig-Holstein, Niedersachsen und Sachsen-Anhalt werden trotz der hohen Anzahl an Grundwassermessstellen nur für sehr wenige Standorte die Konzentrationsklasse 2 überschätzt.

Abbildung 92: Tatsächliche und vorhergesagte Konzentrationsklasse exklusiv für moderat mobile Wirkstoffgruppen aller Grundwassermessstellen.



Quelle: eigene Darstellung BGR.

Abbildung 93: Räumliche Verteilung der korrekt vorhergesagten Grundwassermessstellen exklusiv für moderat mobile Wirkstoffe für die Konzentrationsklasse 2 ($> 0,01 \mu\text{g L}^{-1}$; mitte), wo das Ergebnis unterschätzt wird (rechts) und wo es überschätzt wird (links).



Quelle: eigene Darstellung BGR.

3.2.4.2 Erweiterte Feature-Auswahl

3.2.4.2.1 Modellgüte

Die Modelle der Wirkstoffgruppen erreichen auch mit der erweiterten Feature-Auswahl vergleichbare Performances trotz deutlich geringerer Messstellen-Anzahlen. Die Graphiken werden hier nicht gezeigt. Mobile und sehr mobile Wirkstoffe erreichen einen Macro-F1-Score von 0,662. Die Precision liegt bei 0,847 und 0,496 jeweils für die Konzentrationsklassen 1 und 2, während der Recall bei 0,868 bzw. 0,447 liegt. Bei den wenig mobilen Wirkstoffen wurde eine sehr ähnliche Modellgüte erreicht: Der Macro-F1-Score liegt bei 0,66, eine Precision von jeweils 0,899 und 0,416 und ein Recall von 0,886 bzw. 0,446 erreicht. Damit liegen bei beiden Modellen die Performances der zweiten Klasse etwas unter Zufall. Das Modell der moderat mobilen Wirkstoffe erreicht wiederum einen etwas besseren Macro-F1-Score von 0,722, sowie eine Precision von jeweils 0,806 und 0,65 und einen Recall von 0,837 bzw. 0,599. Somit zeigt das Modell der moderaten Wirkstoffe auch für Konzentrationsklasse 2 akzeptable Ergebnisse, worauf in den nachfolgenden Auswertungen der Feature-Importance und den Feature-Effekten eingegangen wird.

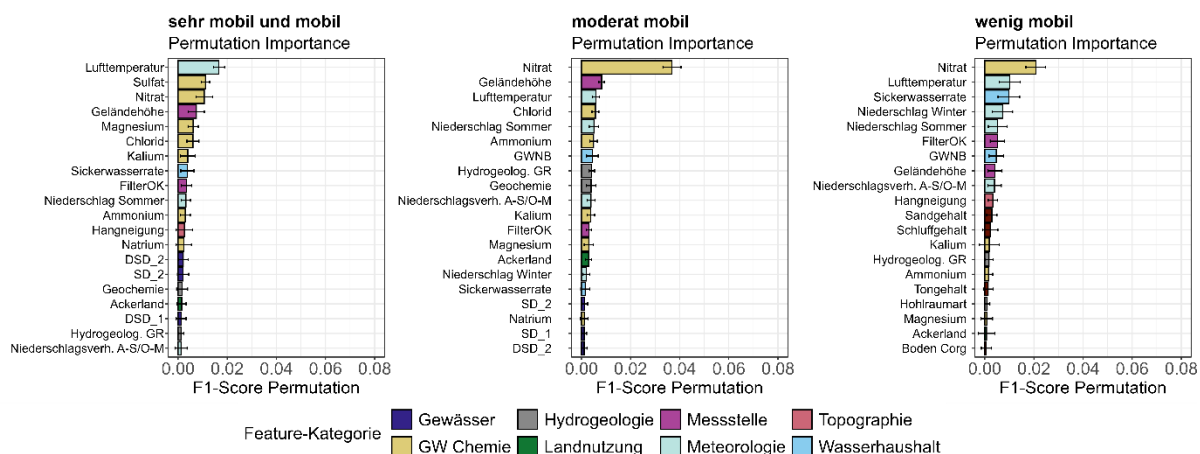
3.2.4.2.2 Feature-Importance

Bei den wenig mobilen Wirkstoffen ist gemäß Permutation Importance (Abbildung 94) der mittlere Nitratgehalt mit signifikantem Abstand das wichtigste Feature. Die Lufttemperatur, Sickerwasserrate, Niederschlag in Winter und Sommer, die Filteroberkante und weitere Features folgen mit nicht klar unterscheidbarem Ranking. Nach Gini impurity (hier nicht gezeigt) sind Nitrat, die Lufttemperatur, das Niederschlagsverhältnis, die Sickerwasserrate, Niederschlag Winter an oberster Stelle, aber in ihrer Abfolge kaum zu unterscheiden. Auch gemäß korrigierter Gini impurity ist der Nitratgehalt gemeinsam mit der Sickerwasserrate, dem Winterniederschlag, dem Niederschlagsverhältnis und der Lufttemperatur als wichtigste Features ausgewiesen.

Bei den mobilen und sehr mobilen Wirkstoffen wird gemäß Permutation Importance die Lufttemperatur als wichtigstes Feature identifiziert, danach folgen der mittlere Sulfatgehalt, Nitratgehalt, die Geländehöhe, Magnesium, Chlorid und Kalium. Nach Gini impurity ist ebenfalls die Lufttemperatur bedeutend, unterscheidet sich aber kaum von der von Nitrat, Sulfat und Chlorid. Danach folgen u.a. Kalium, Natrium, die Sickerwasserrate und die Geländehöhe. Zahlreiche nachfolgende Features zeigen ähnlich hohe Importance Werte auf. Auch korrigierte Gini impurity zeigt die Lufttemperatur an erster Stelle, danach werden Sulfat, Chlorid, Nitrat gelistet.

Die Rangfolge der Feature Importance bei den moderat mobilen Wirkstoffen wird gemäß Permutation Importance von dem mittleren Nitratgehalt angeführt. Mit etwas Abstand folgen die Geländehöhe, die Lufttemperatur, Chlorid, der Sommerniederschlag und weitere Features. Gemäß Gini impurity zeigt sich ein ähnliches Bild mit Nitrat an der Spitze, in den nachfolgenden Features reiht sich wiederum auch der Ammoniumgehalt ein. Nitrat ist auch nach korrigierter Gini impurity das wichtigste Feature. Nachfolgend kommt hier wiederum auch die Aquifer-Geochemie hinzu.

Abbildung 94: Feature-Importance gemäß Permutation für die Wirkstoffgruppen nach Mobilität (erweiterte Feature-Auswahl).



Quelle: eigene Darstellung BGR.

3.2.4.2.3 Feature-Effekte

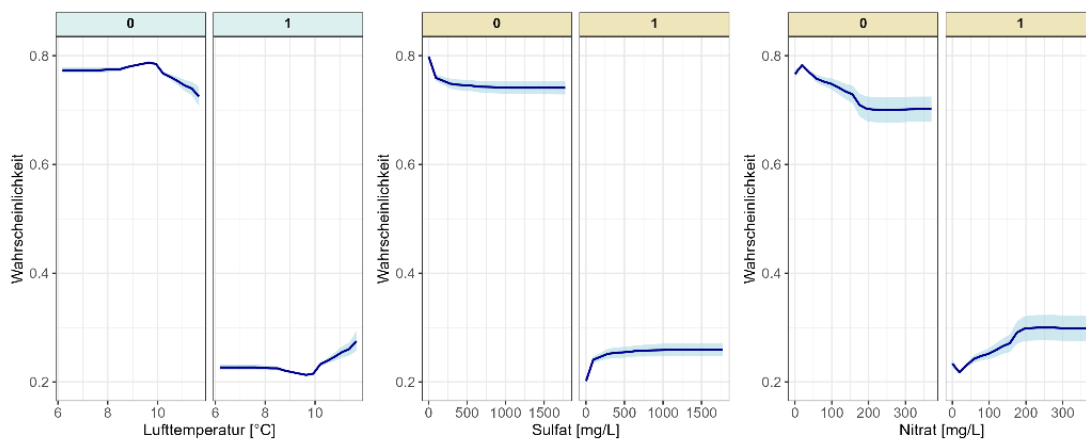
Die Feature-Effekte der jeweils drei wichtigsten Features (gemäß Permutation) des Modells der mobilen und sehr mobilen Wirkstoffe ist in Abbildung 95 dargestellt. Bis zu einer durchschnittlichen langjährigen Lufttemperatur von etwa 10 °C verhält sich die Effektkurve relativ flach und zeigt danach einen Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 mit zunehmender Temperatur. Beim mittleren Sulfatgehalt zeigt sich ein anfänglicher Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 mit zunehmender Konzentration, bei höheren Werten flacht die Kurve ab. Auch Nitrat zeigt einen Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 mit zunehmender Konzentration. Eine Abnahme der Geländehöhe ist von 50 bis 100 m mit einer Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 assoziiert. In größeren Höhen zeigt sich kein Effekt mehr.

Bei den moderat mobilen Wirkstoffen zeigt sich ein anfänglich starker Anstieg der Wahrscheinlichkeit der Konzentrationsklasse 2 bei steigendem Nitratgehalt. Dies wird auch für den Chloridgehalt deutlich. Ein Anstieg der Geländehöhe ist bis etwa 500 m mit einer Zunahme der Wahrscheinlichkeit der Konzentrationsklasse 2 assoziiert. Die Effektkurve der

Lufttemperatur entspricht grundsätzlich dem für die mobilen Wirkstoffe sichtbaren Effekt (Abbildung 96).

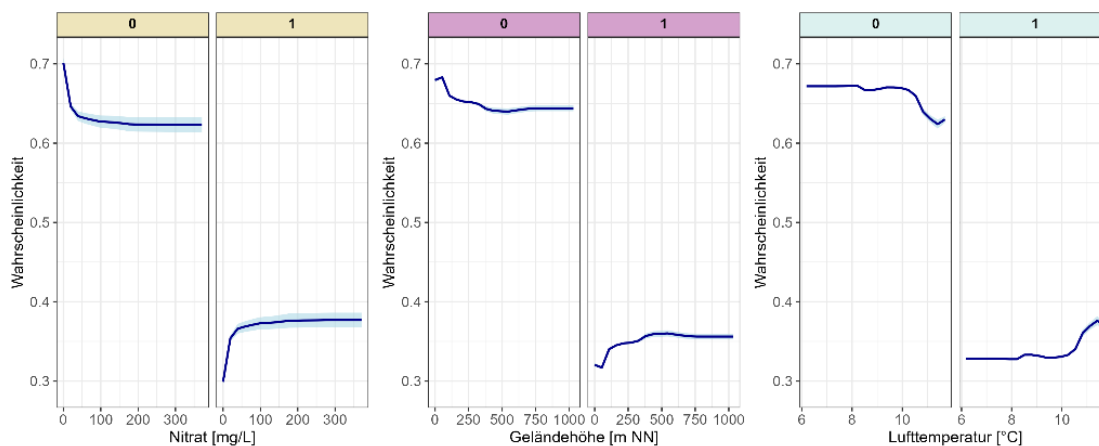
Bei den wenig mobilen Wirkstoffen sind selbst die Effektkurven der drei wichtigsten Features relativ flach. Die grundsätzliche Richtung der Effekte von Nitrat und Lufttemperatur ähneln den bereits genannten Effekten bei den anderen Wirkstoffgruppen. Zudem ist ein leichter negativer Zusammenhang mit der Sickerwasserrate und der Wahrscheinlichkeit der Konzentrationsklasse 2 angezeigt (Abbildung 97).

Abbildung 95: Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der mobilen und sehr mobilen Wirkstoffe mit erweiterter Feature-Auswahl.



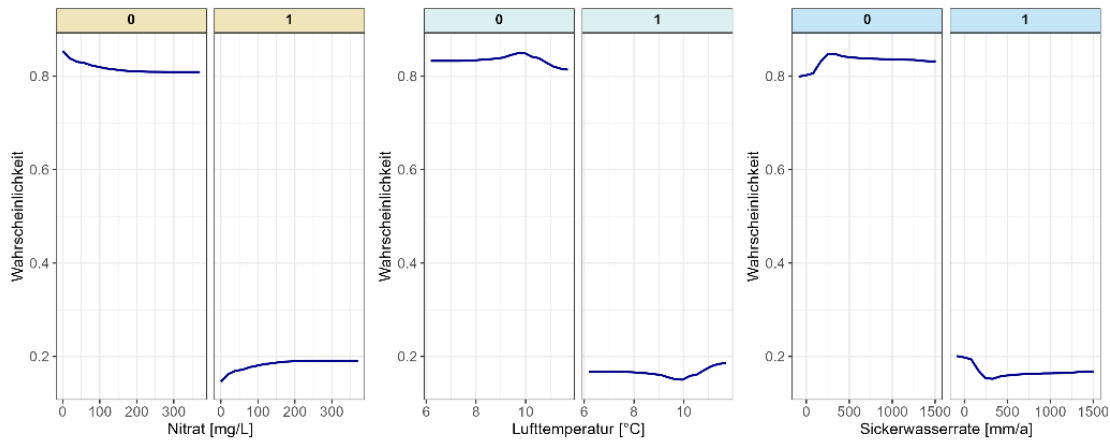
Quelle: eigene Darstellung BGR.

Abbildung 96: Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der moderat mobilen Wirkstoffe mit erweiterter Feature-Auswahl.



Quelle: eigene Darstellung BGR.

Abbildung 97: Partial-Dependence-plot der drei wichtigsten Features (gemäß Permutation Importance) des binären Modells der wenig mobilen Wirkstoffe mit erweiterter Feature-Auswahl.



Quelle: eigene Darstellung BGR.

3.2.4.3 Zusammenfassung

Die Vorhersagegüten der mobilen und sehr mobilen bzw. wenig mobilen Wirkstoffgruppenmodelle (reduzierte Feature-Auswahl) sind hinsichtlich der Vorhersage der höheren Konzentrationsklasse nicht robust genug, daher ist die Feature-Importance dieser Modelle mit Unsicherheiten behaftet. Die Modelle der mobilen und sehr mobilen Wirkstoffe zeigen jedoch eine Feature-Importance auf, die der der Metaboliten ähnelt. Die moderat mobilen Wirkstoffe weisen den hydrogeologischen und Boden-Features eine höhere Gewichtung zu als dem Feature Ackerland. Bei den Modellen der erweiterter Feature-Auswahl wurde bei allen Stoffgruppen und den Einzelstoffen eine leicht bessere Modellgüte erreicht. Der Nitratgehalt ist wichtigstes Feature bei den moderat mobilen und wenig mobilen Wirkstoffen.

4 Diskussion

Da eine Vielzahl an Modellversionen gerechnet wurde, folgt nun eine Übersicht und kritische Diskussion der wichtigsten Erkenntnisse aus den Ergebnissen.

4.1 Metaboliten-Modelle

4.1.1 Reduzierte Feature-Auswahl

4.1.1.1 Konzentrationsklassen und Vorhersagegüte

Die binären Metaboliten-Modelle der reduzierten Feature-Auswahl zeigen (mit Ausnahme von Metazachlorsäure) einen Macro-F1-Score von über 0,7 und die Mehrheit der Messstellen (~80%) konnten korrekt vorhergesagt werden. Somit konnten mit einer überschaubaren Anzahl von räumlichen und zeitlich-aggregierten Features räumliche Modelle für acht Metaboliten mit guter Vorhersagegüte trainiert werden (siehe auch Cooke et al., 2025).

Es wurde sich primär auf die binären Modelle konzentriert. Die Klassengrenze von $0,05 \mu\text{g L}^{-1}$ war sinnvoll gewählt, da damit akzeptable Modellgüten erreicht sowie gegensätzliche Effekte in den beiden Klassen abgedeckt werden konnten. Die Ergebnisse der 3-Klassen-Modelle zeigen, dass die dritte Klasse bezüglich der Interpretation der Feature-Effekte keinen Mehrwert gegenüber den binären Modellen ergibt. Zudem konnte die dritte Klasse aufgrund der geringen Anzahl an Messstellen, die in diese Klasse fallen, bei allen Stoffen nicht mit ausreichender Güte vorhergesagt werden. Die Gütemaße Precision und der Recall waren oftmals für die dritte Klasse schlechter als Zufall (kleiner als 0.33). Aber auch bei den binären Modellen zeigte sich der negative Effekt der Unausgeglichenheit der Anzahl der Messstellen in den zwei Konzentrationsklassen. Stoffe mit stärker ausgeglichener Messstellenverteilung zeigten höhere Precision und Recall für die zweite Konzentrationsklasse, was sich insgesamt auf den Macro-F1-Score auswirkte. Bei den binären Modellen der reduzierten Feature-Auswahl ist die Precision der zweiten Klasse bei nicht allen Stoffen optimal (bei Metazachlorsäure unter 0,5). Da TFA in zum Teil deutlich höheren Konzentrationen vorkommt, stellte sich eine Klassengrenze von $0,5 \mu\text{g L}^{-1}$ als geeigneter heraus. Aufgrund einer gewissen „Hintergrundkonzentration“ durch zahlreiche zusätzliche Eintragspfade wie z.B. atmosphärische Deposition war es dem Modell anhand der gegebenen Features nicht möglich, Unterschiede zwischen den beiden ursprünglichen Klassen zu erlernen (Albers et al. 2024). Zudem wirkte sich das größere Ungleichgewicht der Messstellenverteilung zwischen den Klassen negativ auf die Modellgüte aus. Mit der höheren Klassengrenze konnte eine Unterscheidung der zwei Klassen gelernt werden.

Bundesweit konnte die Mehrheit der Messstellen korrekt vorhergesagt werden. In einzelnen Bundesländern und kleinräumigeren, naturräumlich und landwirtschaftlich abgrenzbaren Gebieten kann sich diese Verteilung verschieben. Die räumliche Verteilung von korrekt vorhergesagten sowie von über- bzw. unterschätzten Messstellen zeigten je nach Stoff vereinzelt Regionen, die zur Über- oder Unterschätzung tendieren. Perspektivisch ist eine Analyse der abweichenden Messstellen mit Hilfe regionaler Kenntnisse von Vorteil, um die Gründe für Abweichungen besser zu verstehen und um die Modelle in diesen Regionen weiterzuentwickeln. Hierzu müssten weitere Features berücksichtigt werden, wie beispielsweise die relative Lage im Einzugsgebiet nach Strahler-Ordnung und damit der potenzielle Eintrag von Oberflächenwasser. Diese Features und weitere, die den Abstand zu Gewässern beschreiben, wurden in den Modellen der erweiterten Feature-Auswahl getestet und zeigten für diese keine Bedeutung. Dies ist vermutlich der Fall, da eine möglicherweise vorhandene zeitliche Dynamik der Grundwasser-Oberflächenwasser-Interaktion nicht berücksichtigt werden konnte. Es ist perspektivisch

dennoch sinnvoll, in den konkreten Gebieten gewisse lokal relevante Features noch einmal aufzugreifen (beispielsweise höher aufgelöste Bodeneingangsdaten), um die Modellresiduen besser zu verstehen. Es ist möglich, dass Features, die bundesweit nicht nennenswert zu den Modellen beitragen, dennoch lokal zur Erklärung von Modellabweichungen herangezogen werden können.

Neuronales Netz und Random-Forest zeigen Effekte mit gleicher Tendenz an, während die Effektkurven beim Neuronalen Netz generell glatter verlaufen. Beide Modellarchitekturen erzielen vergleichbare Vorhersagegüten, wobei die Random-Forest-Modelle etwas höhere erzielen. Modelle beider Architekturen zeigen eindeutig die große Bedeutung der Anbaufläche der jeweiligen Hauptkultur im Puffer bzw. der Tiefe der Filteroberkante der Messstellen als erklärende Variablen sowie die Zusammenhänge innerhalb der Konzentrationsklassen.

4.1.1.2 Feature-Gewichtung und Feature-Effekte der reduzierten Feature-Auswahl

Bei den Modellen mit reduzierter Feature-Auswahl und für Metaboliten, deren Muttersubstanzen in einer überschaubaren Anzahl an Kulturen angewendet werden (alle außer TFA), wurde jeweils der Anbau der Hauptkultur mit Abstand als wichtigstes Feature identifiziert. Dieses Ergebnis ist unabhängig vom verwendeten Algorithmus (Random-Forest oder Neuronales Netz) und der Feature-Importance-Methode (Permutation oder Gini impurity). Die Landnutzungs-Features beruhen auf der Mittelung der Flächenanteile einer Kultur in den Jahren 2017-2019. Somit beinhaltet dieses Feature keine Information über vorherige Anbaumuster und drei Jahre sind für die Ermittlung von Fruchtfolgen z.T. zu kurz. Dass die Hauptkultur dennoch als das wichtigste Feature für die meisten Metaboliten identifiziert wurde, ist vermutlich damit zu erklären, dass die Jahre 2017-2019 grundsätzliche, großräumige und langjährige Anbaumuster nach wie vor gut abdecken. Die Nebenkulturen zeigen eine geringere Gewichtung als die Hauptkultur. Zwar wurden Kulturen berücksichtigt, für die eine Zulassung besteht bzw. bestand. Dennoch können sich die Zusammenhänge im Modell auch aus der räumlichen Korrelation zwischen den Kulturen am Standort und aufgrund von Fruchtfolgen erklären. Der Anteil der Anbaufläche der Hauptkultur zeigt für alle Stoffe einen positiven Zusammenhang mit relativ steiler Effektkurve. Nebenkulturen zeigen den gleichen Trend, oft in abgeschwächter Form. Bei TFA, einen Metaboliten, der aus mehreren PSM-Wirkstoffen entstehen kann und darüber hinaus noch andere Eintragsursachen als die landwirtschaftliche Nutzung hat, zeigt sich, dass der Einfluss nicht auf den Anbau einzelner Kulturen zurückgeht, sondern die Landnutzung in der Summe betrachtet werden muss. Hierbei kann es von Nachteil sein, dass so viele verschiedene Kulturen als Features mit aufgenommen wurden, da sich Unterschiede zwischen den Kulturen hinsichtlich Feature-Gewichtung und Feature-Effekte kaum herausarbeiten lassen. Dass das TFA-Modell mit den vorhandenen Landnutzungs-Features dennoch eine gute Modellgüte erreicht, ist ein Hinweis darauf, dass sich der Einfluss der Landwirtschaft als Eintragspfad für Grundwasserfunde von TFA abzeichnet und von den Modellen erlernt werden konnte. Dies zeigt sich u.a. daran, dass in der Rangordnung der Gewichtung die Wintergerste vorkommt, der die Anwendung von Flufenacet zugeordnet werden kann.

Nachfolgende Features im Ranking stellen die Filteroberkante, der Sandgehalt als Vertreter für die Korngrößenverteilung im Boden, die Grundwasserneubildung und der organische Kohlenstoff im Boden dar, deren Rankings sich je nach Metaboliten, Algorithmus und Methode unterscheiden können. Je näher der Filter einer Messstelle an der Oberfläche verbaut ist, desto größer ist die Wahrscheinlichkeit für höhere Konzentrationen (repräsentiert durch Konzentrationsklasse 2) für alle untersuchten Stoffe. Der Sandgehalt, stellvertretend als eines der drei Wertepaare (i.e., Sand, Schluff, Ton) zur Ableitung der Bodenart, wurde meist als das wichtigste Boden-Feature eingeordnet. Hinsichtlich der Feature-Effekte sind Unterschiede

zwischen den Stoffen beim Sandgehalt, der Grundwasserneubildung und dem organischen Kohlenstoff im Boden sichtbar.

Es ist nicht trivial, tatsächliche Feature-Effekte besser von statistischen Effekten abzugrenzen, die aus Korrelationen mit dem wichtigen Feature der jeweiligen Hauptkultur resultieren. Zur Visualisierung wurden die Verteilungen der Features an allen Messstellen mit jenen an Messstellen in Anbaugebieten verglichen und im Kontext der Feature-Effekte diskutiert. Zeigt ein Feature in Anbaugebieten Ausprägungen, die deutlich von der mittleren Feature-Ausprägung aller Messstellen-Puffer abweicht, kann nicht ausgeschlossen werden, dass die spezifische Feature-Ausprägung als Indikator für Anbaugebiete und somit für eine erhöhte Anwendung von bestimmten PSM dient, wie dies beispielsweise beim Sandgehalt bei manchen Stoffen zu sehen ist. Der Sandgehalt als auch die zugehörige Anbaufläche weisen beide eine hohe Permutation Importance auf, sind jedoch nur schwach korreliert und zeigen geringe Feature Interactions. Daraus lässt sich ableiten, dass beide Feature nicht zwangsläufig als redundant, sondern als komplementär zu verstehen sind, da beide Features sich jeweils ergänzend zur Identifizierung von Anbauaktivitäten und damit von der Anwendung von Pflanzenschutzmitteln beitragen können.

Die stärksten Feature-Effekte für Sand wurden für die Chloridazon-Metabolite und für TFA deutlich. Höhere Konzentrationen der Chloridazon-Metaboliten im Grundwasser treten bei geringeren und abnehmenden Sandgehalten und dadurch zwangsläufig bei höheren und zunehmenden Schluff- und Tongehalten der Böden auf. Gleichzeitig zeigt der Vergleich der Mediane einen geringeren Sandgehalt der Böden für alle Messstellen gegenüber den Messstellen in den vorrangigen Anbaugebieten der Zuckerrübe. Es ist möglich, dass ein geringer Sandgehalt (höherer Anteil an Schluff und Ton) auch als Proxy für Zuckerrüben-Anbaugebiete verstanden werden kann bzw. dass beide Informationen in Kombination auf den PSM-Einsatz hindeuten. Es ist nicht möglich, auf direkte kausale Zusammenhänge zwischen der Korngrößenverteilung der Böden für die Vorhersage von höheren Funden der Chloridazon-Metaboliten im Grundwasser zu schließen. Vielmehr zeigt die Analyse, dass Grundwasserfunde der beiden untersuchten Abbauprodukte von Chloridazon über das gesamte Spektrum der Korngrößenverteilung typischer Böden in Zuckerrübenanbaugebieten vorkommen. Auch für die untersuchten Metazachlor-Metaboliten liegen in Anbaugebieten in der Tendenz geringere Sandgehalte vor. Andererseits zeigen die Feature-Effekte jedoch einen positiven Zusammenhang zwischen Sandgehalt und höheren Stoffkonzentrationen. Für die Metolachlor-Metaboliten sind dagegen keine starken Unterschiede in der Verteilung des Sandgehalts zwischen den Anbaugebieten und allen Messstellen ersichtlich. Die Feature-Effekte hatten jedoch einen leicht positiven Zusammenhang zwischen Sandgehalt und höheren Konzentrationen aufgezeigt. Insgesamt heißt das, dass für einen Teil der Stoffe (Chloridazon-Metaboliten) die Feature-Effekte der Verteilung der Bodenarten in den Anbaugebieten zu folgen scheinen, während dies für andere Stoffe nicht eindeutig ist. Grundsätzlich muss berücksichtigt werden, dass der Sandgehalt und die Korngrößenzusammensetzung der Böden eine statische Größe ist, die sich über die Zeit nicht ändert (Weil et al. 2017). Der Anbau diverser Hauptkulturen gemäß einer Fruchtfolge ist dagegen zeitlich sehr variabel und auch räumlich (durch die administrativen Begrenzungen des jeweiligen Feldes) sehr variabel. Ursache-Wirkungsbeziehungen zwischen diesen Größen Sandgehalt und Hauptkultur sind also auch unter dem räumlich-zeitlichen Maßstab nicht unisono durchzuführen.

Der Einfluss der Makroporenklassen muss im Zusammenhang mit den Korngrößenverteilungen der Böden interpretiert werden. Beide Features stehen in einem starken Zusammenhang miteinander, da die Makroporenklassen hauptsächlich auf den Ton- und Schluffgehalten der Böden basiert. Die Feature-Effekte des Sandgehaltes und der Makroporenklassen verhalten sich

gegenläufig. Dies wird bei den Chloridazon-Metaboliten deutlich, dass höhere Konzentrationen in Verbindung zu abnehmenden Sandgehalten (zunehmenden Schluff-/Tongehalten) stehen. Gleichzeitig zeigt sich bei diesen beiden Stoffen ein leichter Zusammenhang zwischen höherer Makroporenklasse und höherer Konzentrationsklasse. Bei Metolachlor ESA zeigt sich dieser Zusammenhang ebenfalls, jedoch umgekehrt. Zunehmende Sandgehalte stehen im Verhältnis zu abnehmenden Makroporenklassen.

Die nationale PSM-Grundwasserbewertung (Lysimeterstudien) sowie die Standortauswahl von NZM-Standorten orientiert sich bisher an der erhöhten Versickerungsneigung an sandigen Standorten. Die betrachteten mobilen Metaboliten kommen auch an Standorten mit bindigen Böden vor, wenn dort die zugehörigen Kulturen angebaut wurden und man vermuten kann, dass die jeweilige Muttersubstanz auch angewendet wurde. Es stellt sich die Frage, weshalb in bindigen Böden mit einer höheren Verweilzeit (längerer Rückhalt im Boden für höhere Abbaurate) dennoch so häufig Funde vorliegen. Hier spielt mit Sicherheit die lange Anwendungshistorie der Stoffe eine Rolle. Zumindest sei aber empfohlen, auch Standorte mit tonigen und lehmigen Böden als NZM-Standorte in Betracht zu ziehen, da ihre Relevanz aus diesen Modellergebnissen zumindest für Metaboliten belegt ist. Ob die Wirkung von Makroporenfluss an Standorten mit schweren Böden aus den Modellergebnissen direkt abzulesen ist, ist diskussionswürdig. Das Feature Sandgehalt und die Makroporenklassen sind über das Bodentexturdreieck zu einem gewissen Grad korreliert, wenn auch nicht komplett invers. Die relativ geringe Feature-Gewichtung der Makroporenklasse im Vergleich zum Sandgehalt kann unter Umständen darin begründet liegen, dass ein kontinuierliches Feature (Sandgehalt) gegenüber eines aus drei Klassen bestehenden kategorischen Features (Makroporen) im Modell leichter zu handhaben ist. Da vieles dafürspricht, dass der Sandgehalt seine Gewichtung vorwiegend aufgrund seiner Funktion als Zeiger für typische Anbaugebiete erhält, lässt sich argumentieren, dass dies auch auf die Makroporenklassen zutrifft. Dies zeigt sich durch die bei einigen Stoffen jeweils gegenläufig verlaufenden Feature-Effekte für Sand und Makroporen. Wir ziehen das Fazit, dass die Makroporenklasse - ähnlich dem Sandgehalt, nur umgekehrt - komplementär zum Feature Anbaufläche auf Gebiete mit PSM-Einsatz hindeutet. Anhand der Ergebnisse kann im Umkehrschluss nicht der Nachweis erbracht werden, dass solche mobilen Metaboliten nach langjähriger Anwendung ihrer Muttersubstanzen verstärkt auf Standorten mit sandigen Böden ins Grundwasser gelangen können.

Das Potenzial der Makroporenbildung bzw. präferenzierter Fließwege anhand von Pedotransferfunktionen räumlich zu schätzen, ist keine triviale Aufgabe. Der hier verwendete Ansatz wurde aus einem Forschungsprojekt (Klein et al. 2019; Klein et al. 2025) übernommen, in welchem drei Makroporenklassen als Grundlage für die Berechnung präferenzierellen Stofftransports in der ungesättigten Zone als zusätzlicher Prozess zum kapazitätsbasierten Matrixfluss in einem eindimensionalen Versickerungsmodell (GeoPELMO-DE) verwendet wurde. In den vorliegenden Analysen wurden die drei Makroporenklassen, die ein Potenzial der Makroporenbildung abbilden, vereinfacht (ohne dahinterliegende Stofftransportflüsse) auf Monitoringdaten aus der gesättigten Grundwasserzone angewendet. Es ist davon auszugehen, dass die verwendete Datengrundlage an zeitlich aggregierten Monitoringdaten sowie vereinfachten Makroporenklassen der räumlichen (und zeitlichen) Dynamik der Ausbildung von Makroporenfluss (präferenzielles Fließen) nicht gerecht wird und ein Zusammenhang somit in den Modellen des Maschinellen Lernens nicht hinreichend abgebildet werden kann. Makroporenfluss steht vor allem für einen kurzfristig schnelleren Transport von Stoffen, der im Jahresgang temporär und zudem nicht flächendeckend, sondern lokal stattfinden kann. Während die Modellentwicklung von GeoPELMO DE zum Ziel hatte, die Versickerungsleistung von schweren Böden gegenüber sandigen Böden durch die zusätzliche Berücksichtigung von präferenziellem Fließen nicht weiter zu unterschätzen, können wir anhand unserer

Projektanalysen nicht schlussfolgern, dass langfristig höhere Grundwasserkonzentrationen anhand der vorliegenden Makroporenklassen erklärt werden können. Dennoch wurde anhand der Projektergebnisse gezeigt, dass die mobilen bis sehr mobilen Metaboliten relativ unabhängig von der Korngrößenzusammensetzung (dem Sandgehalt) der Böden vorkommen. In dem Projekt kann nicht abschließend geklärt werden, ob sich darin ein Hinweis auf präferenzielle Fließwege auf bindigen Böden verbirgt oder ob allein Matrixfluss die Ursache dafür ist. Allerdings stehen die Ergebnisse in einem gewissen Widerspruch zu den Annahmen in der nationalen Risikobewertung für PSM, wo sandige Böden als „realistic worst case“-Szenarien für die Versickerung angesehen werden. Dabei ist aber auch zu beachten, dass zeitlich gemittelte Daten nach einigen Jahrzehnten der Ausbringung von Altstoffen möglicherweise nicht herangezogen werden können, um Aussagen zum Einfluss präferenzierter Fließwege zu treffen. Weiterhin ist zu diskutieren, inwiefern die Makroporenklassen, die bis zu 1 m Tiefe abgeleitet wurden, sich auf den oftmals erheblich tieferen Grundwasserspiegel auswirken. Präferenzielle Fließwege können existieren und zeitlich und räumlich begrenzt einen Einfluss auf den Stofftransport zeigen, aber ob dieser Effekt auch langfristig auf Landschaftsebene (Einzugsgebietsebene) nachweisbar ist, ist nicht zwangsläufig gegeben. Zusammenfassend lässt sich sagen, dass das Signal von Makroporenfluss in zeitlich stark aggregierten Konzentrationsklassen der Metaboliten, wie das im Projekt erfolgte, nicht aufgelöst werden kann. Somit ist davon auszugehen, dass anhand dieser Daten und der entwickelten Modelle des Maschinellen Lernens keine abschließenden Schlussfolgerungen hinsichtlich der Relevanz von Makroporenfluss für die Belastung von Messstellen mit den untersuchten Metaboliten getroffen werden können.

Mit Ausnahme der Metolachlor-Metaboliten zeigten alle Metaboliten einen geringeren Zusammenhang mit der Grundwasserneubildung in den Anbaugebieten an. Bei den Metolachlor-Metaboliten zeigte der Feature-Effekt wiederum einen positiven Zusammenhang zwischen der Grundwasserneubildung und der höheren Konzentrationsklasse, was sich mit der höheren Grundwasserneubildung in den Anbaugebieten deckt. Bei den Chloridazon-Metaboliten ist dieser Zusammenhang nicht zu sehen. Bei Metazachlor ESA und der Metazachlorsäure ist wiederum ein negativer Zusammenhang zwischen der Grundwasserneubildung und der höheren Konzentrationsklasse beobachtet worden. Zusammenfassend lässt sich sagen, dass der Einfluss der Grundwasserneubildung auf die Konzentrationen sich bei manchen Stoffen (Metolachlor-Metaboliten) mit einer höheren Grundwasserneubildung in den Anbaugebieten deckt, während dies bei den anderen Stoffen nicht ersichtlich wird.

Ein negativer Zusammenhang zwischen der Tiefe der Filteroberkante und der Wahrscheinlichkeit für Konzentrationsklasse 2 wurde für alle Stoffe gefunden. Das heißt, je höher der Filter einer Messstelle an der Oberfläche verbaut ist, desto größer ist die Wahrscheinlichkeit für höhere Konzentrationen der untersuchten 8 Metaboliten im Grundwasser. Unterschiede zu Messstellen innerhalb der Anbaugebiete wurden nicht deutlich. Dies spricht dafür, dass ein grundsätzlicher Zusammenhang gelernt worden ist, der nicht auf räumlichen Korrelationen der Feature-Ausprägungen mit der jeweiligen Hauptkultur beruht.

Die Verteilung der Wertebereiche des organischen Kohlenstoffs im Boden unterscheidet sich wenig zwischen allen im Datensatz untersuchten Messstellen und den Messstellen innerhalb der Anbaugebiete. Die Feature-Effekte der Random-Forest-Modelle zeigen bei den Stoffen z.T. unterschiedliche Tendenzen, die bei den Modellen, die mit Neuronalen Netzen trainiert wurden, deutlich flacher und glatter ausfallen. Problematisch für die Interpretation der Modellergebnisse in diesem Fall ist die Tatsache, dass ein höherer Sandgehalt meist mit einem geringeren C_{org} -Gehalt einhergeht. Somit sind die Feature-Effekte sehr wahrscheinlich z.T. auch als Zeiger für landwirtschaftliche Gebiete zu sehen. Ein weiterer wichtiger Faktor ist die hohe Mobilität der

Stoffe und die langen Anwendungszeiträume, sodass der Effekt des Rückhaltevermögens an der organischen Substanz aus zeitlich gemittelten Grundwasserdaten schwer extrahiert werden kann. Kausale Zusammenhänge zwischen dem Rückhaltevermögen der organischen Substanz und PSM-Wirkstoffen und deren Metaboliten lässt sich zum jetzigen Stand nur aus kleinskaligen Laborversuchen ermitteln. Dabei werden Bodensäulen mit unterschiedlichen C_{org} -Gehalten manipuliert, bei ansonsten identischen Bedingungen. Durch Batch-Versuche oder Beregnung der Säulen und Analyse des aufgefangenen Perkolats lassen sich anhand von Durchbruchkurven eines Wirkstoffs die Retentionseigenschaften des Bodens ableiten.

Die gezeigten Analysen sind ein Versuch, mögliche Zusammenhänge zwischen Featurewerten und Anbaugebieten darzustellen. Daraus kann man folgern, dass ein Erlernen der räumlichen Zusammenhänge zwischen Anbaugebieten und Features bei einem Teil der Stoffe und Features naheliegend sein kann. Es zeichnet sich jedoch ab, dass die Features tendenziell eher komplementär zur räumlichen Identifizierung von PSM-Einsatz beitragen und nicht bloß auf räumlichen Kreuzkorrelationen zwischen dem Feature und den Anbaugebieten beruhen. Zusätzlich kann die räumliche Verteilung der jeweils in den Modellen einbezogenen Messstellen dazu führen, dass anstatt relevanter Zusammenhänge bzgl. des Stofftransports Artefakte gelernt werden, die wiederum zu Fehleinschätzungen führen.

Die Modelle basieren hauptsächlich darauf, dass eine größere Anbaufläche der relevanten Kulturen im Puffer einer Messstelle sowie eine näher an der Oberfläche gelegene Filteroberkante mit einer höheren Wahrscheinlichkeit für die höhere Konzentrationsklasse assoziiert sind. Die Analyse der Feature-Interaktionen hat zudem gezeigt, dass die Modelle auf vorwiegend leichten bis moderaten Interaktionen basieren, was für ein additives Zusammenwirken der Features spricht. D.h., der Einfluss eines Features verändert sich nicht durch die Interaktion mit einem anderen Feature. Das deutet darauf hin, dass die Features weitgehend unabhängig voneinander wirken und ihre Effekte sich näherungsweise additiv zum Gesamtergebnis zusammensetzen könnten. Anhand der Residuen (*Differenz zwischen den tatsächlichen beobachteten Werten und den vorhergesagten Werten einer statistischen Analyse*) wurde deutlich, dass Messstellen, an denen es zu Überschätzungen der Grundwasserkonzentrationen kommt, tendenziell etwas geringere Tiefen der Filteroberkante sowie etwas größere Flächenangaben der korrespondierenden Hauptkultur aufweisen. Auch wenn die Unterschiede z.T. gering sind, lässt sich daraus schlussfolgern, dass das Modell an denjenigen Messstellen zu einer Fehleinschätzung kommt, an denen der zuvor beschriebene, simple Zusammenhang zwischen Anbaufläche der Kultur und der Tiefe der Filteroberkante nicht ausreicht, um die Konzentrationsklasse korrekt zu schätzen. Die wichtigste Information, wann und in welcher Menge die PSM-Wirkstoffe in der Vergangenheit tatsächlich ausgebracht wurden, ist nicht vorhanden. Sie könnte eine entscheidende Rolle spielen, um Über- bzw. Unterschätzungen an Messstellen zu erklären. Das von den Modellen erlernte, vereinfachte Zusammenspiel aus Anbaufläche und Filteroberkante ist an einem Teil der Messstellen unzureichend, da der Faktor der tatsächlich aufgebrauchten Stoffmengen fehlt. Die fehlenden Anwendungsdaten sind nicht die einzige mögliche Erklärung. An einem Teil der Messstellen kann die potenzielle Unsicherheit im Modell durch die Verwendung der Filteroberkante als Proxy für fehlende bundesweite Grundwasserflurabstände in diesem Zusammenhang relevant sein. Hier ist zu erwähnen, dass auch die Information zum Grundwasserflurabstand fehlerbehaftet sein kann, wenn dieser auf sehr wenigen Messungen basiert und in der Vergangenheit weit zurückreicht. Die Unsicherheit der Filteroberkante lässt sich daher bei zahlreichen Messstellen aufgrund fehlender Informationen zum Grundwasserflurabstand nicht robust quantifizieren. Ein weiterer Unsicherheitsfaktor ist das zeitliche Aggregieren von Landnutzungsinformationen über nur drei Jahre, weshalb an einem Teil der Messstellen die Flächenangaben zu den Kulturen die langfristige Landnutzung nicht immer korrekt abbilden.

Die hydraulische Leitfähigkeitsklasse des Aquifers, die Hohlraumart sowie die Makroporenklasse wurden bei allen Metaboliten und angewendeten Methoden durchweg als weniger relevant eingestuft. Auch zeigen sich hier eher schwache Feature-Effekte. Die sehr ungleich verteilten Messstellenanzahlen in den Klassen der hydraulischen Leitfähigkeit, der Makroporen und auch der Hohlraumart könnten dazu beigetragen haben, dass aus diesen Features keine nennenswerten Effekte gelernt werden konnten. Bei einem Teil der Stoffe zeigt die Hohlraumart Karst auffällige Feature-Effekte, wobei sich hierbei nicht klar sagen lässt, ob diese durch die Verteilung der Messstellen in diesen Gebieten oder auch die Verteilung der Hohlraumart in Anbaugebieten erklärbar sind. In Kapitel 5 wird vertieft darauf eingegangen, ob eine hohe Mobilität der Metaboliten eine Rolle spielen könnte, weshalb der Einfluss der Aquifer-Eigenschaften in den Modellen derart gering ausfällt.

4.1.2 Erweiterte Feature-Auswahl

Die Modelle der erweiterten Feature-Auswahl zeigen eine höhere Modellgüte als die Modelle der reduzierten Feature-Auswahl, trotz einer geringeren Anzahl an Messstellen. Die zusätzlichen Features konnten entscheidend dazu beitragen, die Vorhersagegüte zu erhöhen.

Die Hauptkultur ist auch in diesen Modellen in den meisten Fällen das wichtigste Feature. Für Metolachlor ESA ist mit ähnlich hoher Feature-Gewichtung der Kaliumgehalt zu nennen, wogegen für TFA der Nitratgehalt das wichtigste Feature ist. Die grundwasserchemischen Features Kalium, Nitrat, Sulfat (für TFA auch Ammonium) und weiterhin auch die Anbaufläche an Acker allgemein im Puffer der Messstelle dominieren das Feature-Ranking. Es ist weiterhin sinnvoll auch die Features, welche eine geringe Feature-Gewichtung zugeordnet wurde, genauer zu beleuchten. Features, die bei der reduzierten Feature-Auswahl als relevant eingestuft wurden, wie die Tiefe der Filteroberkante, die Grundwasserneubildung, oder auch Boden-Features, werden bei den meisten Stoffen als deutlich weniger bedeutend eingestuft. Features, welche die relative Lage im Einzugsgebiet bzw. den Abstand zu Fließgewässern beschreiben (LP, SD, DSD) oder auch der Hangneigung wird bei den meisten Stoffen nur eine geringe Gewichtung beigemessen.

Die grundwasserchemischen Features zeigten in ihren Effektkurven ein Verhalten, das mit den Feature-Effekten der Landnutzung vergleichbar ist. Zum Teil waren die Effektkurven recht steil, sodass der Effekt eher davon abhängt, ob überhaupt eine Mindestkonzentration im Grundwasser vorkommt oder nicht. Bei höheren Konzentrationen zeigte sich keine Veränderung der Effekte. Der Abgleich mit der Verteilung der Features in den Anbaugebieten ist auch bei der erweiterten Feature-Auswahl aufschlussreich: Bei allen Metaboliten ist der Nitratgehalt im Grundwasser in Anbaugebieten höher. Ausschließlich für TFA zeigen sich keine Unterschiede. Allerdings sind diese Analysen bei TFA nur begrenzt aussagekräftig, da man bei der Vielzahl von Kulturen sowie weiteren Eintragspfaden nicht von typischen Anbaugebieten sprechen kann. Der Kaliumgehalt in den Anbaugebieten ist bei den Chloridazon-Metaboliten gleichbleibend, bei TFA leicht geringer. Bei den Metolachlor- und Metazachlor-Metaboliten ist der Kaliumgehalt in den Anbaugebieten höher. Die Sulfatgehalte sind in den Anbaugebieten bei allen Metaboliten höher außer bei den Metolachlor-Metaboliten, wo diese geringer sind. Ammoniumgehalte sind bei allen Stoffen in den Anbaugebieten leicht höher. Damit zeigt sich recht deutlich, dass die grundwasserchemischen Features (allen voran Nitrat) mit dem Anbau der Hauptkultur in einem räumlichen Zusammenhang stehen. In Kombination mit der Analyse der Feature-Effekte lässt sich schlussfolgern, dass diese grundwasserchemischen Parameter als Zeiger für landwirtschaftliche Aktivität und ggf. als ein Hinweis auf die Intensität dieser gesehen werden können (Davis et al. 2012). Zwar kann das Vorkommen von Nitrat auch ein Hinweis auf ein erhöhtes Versickerungspotenzial sein (Davis et al. 2012; Vinten et al. 1994), in diesem Kontext

ist jedoch davon auszugehen, dass die Modelle Zusammenhänge vorwiegend aufgrund der räumlichen Assoziation lernen. Somit verfeinert die Information über die Nitratgehalte höchstwahrscheinlich die Lokalisation von intensiver Landwirtschaft, was der Modellgüte zugutekommt. Es ist im Kontext der Datenlage und der beschriebenen Modelle jedoch nicht sinnvoll, aus dem mittleren Nitratgehalt (dies betrifft auch Kalium, Sulfat und weitere) auf ein unmittelbares Prozessverständnis im Sinne von Stofftransportprozessen schließen zu wollen. Die über Jahre gemittelten Nitratgehalte können keine zeitliche Dynamik auflösen wie die von z.B. Redoxprozessen. Somit handelt es sich um langfristige Standorteigenschaften, die weniger den Transport und Abbau von Stoffen als vielmehr das Vorhandensein von Landwirtschaft darstellen. Weitere wichtige chemische Parameter, die den Abbau und die Mobilität von PSM-Wirkstoffen beeinflussen, konnten aufgrund von Datenlücken nicht untersucht werden (bspw. Temperatur, pH-Wert, Redoxpotential, Sauerstoff- und Eisenkonzentration). Der Einfluss dieser Features, sofern zeitlich aufgelöst in ausreichender Datenqualität vorhanden, könnte in einem räumlich-zeitlichen Modell erforscht werden und für zukünftige Modelle einen Mehrwert bieten.

Analog zu den grundwasserchemischen Features sind auch die meteorologischen Features zu interpretieren, die zum einen aufgrund ihrer Wechselbeziehung (Multi-Kollinearität) die relative Abfolge der Rangordnung der Feature-Gewichtung verzerren und zum anderen ebenfalls auf landwirtschaftliche Gebiete verweisen können. Features, die langfristige meteorologische Bedingungen darstellen, wie die Lufttemperatur oder der Niederschlag, sind ebenfalls als mögliche Zeiger für Standortfaktoren der Anbaugelände zu bewerten. Genauso ist die Höhenlage oder der hydrogeologische Großraum zu werten.

Insgesamt lässt sich feststellen, dass die Vorhersagegüte der Modelle der erweiterten Feature-Auswahl höher ist, die erhöhte Anzahl an Features allerdings die Interpretation erschweren kann, vor allem, da ein Teil der Features Korrelationen untereinander aufweisen. Somit konnten zwar performantere Modelle trainiert werden, der zusätzliche Erkenntnisgewinn hinsichtlich vulnerabler Gebiete aus zahlreichen Features hält sich jedoch in Grenzen.

Grundwasserchemische Features wie Nitrat und Kalium stechen aufgrund ihres großen Beitrags zur Vorhersage heraus und könnten zusätzlich zu Landnutzungsfeatures als Proxy für PSM-Aufwendungen verwendet werden.

4.2 Wirkstoff- und Wirkstoffgruppenmodelle

Modelle für drei Wirkstoffgruppen (sehr mobile und mobile, moderat mobile sowie wenig mobile Wirkstoffe) und Einzelstoffmodelle wurden für die reduzierte und die erweiterte Feature-Auswahl gerechnet.

Zuerst werden die Ergebnisse der reduzierten Feature-Auswahl diskutiert.

Die Modellgüten der Modelle für die sehr mobilen und mobilen als auch für die wenig mobilen Wirkstoffe waren besonders in Hinblick auf Konzentrationsklasse 2 nicht ausreichend. Ebenso verhielten sich die Metriken der Einzelstoffmodelle für Bentazon und Simazin.

Vielversprechender in Bezug auf die Modellgüte waren wiederum die Modelle der moderat mobilen Wirkstoffe und das Einzelstoffmodell für Atrazin. Grundsätzlich ist zu kritisieren, dass die Wirkstoffgruppenmodelle aufgrund der Vielzahl der Stoffe und der unterschiedlichen Kulturzuordnung der Stoffe und deren Anwendungshistorien und letztendlich auch Messprogrammen verschiedene Blaupausen in den Daten erlernen. Diese Kritik gilt in geringerem Maße auch für die Metaboliten-Einzelstoffmodelle, für die mehrere Kulturzuordnungen greifen. Anstatt einen Großteil der Kulturen als Einzel-Features zu verwenden, wurde bei den Wirkstoffen nur das übergreifende Feature Ackerland untersucht. Das Feature Ackerland im Puffer einer Messstelle zeigt im Ergebnis jedoch in der Gewichtung

und den Feature-Effekten nur einen geringen Einfluss. Somit kann bei Atrazin als ein Altstoff, dessen Anwendung bereits seit Anfang der Neunziger Jahre nicht mehr zugelassen ist, die aktuelle Landnutzung nur sehr bedingt herangezogen werden. Atrazin fließt als moderat mobiler Stoff auch in die Modelle der jeweiligen Wirkstoffgruppe ein. Was die Anzahl der Messstellen anbelangt, die alleinig aufgrund von Atrazin in Konzentrationsklasse 2 fallen, wird das Modell jedoch nicht von Atrazin dominiert, sondern verteilt sich auf zahlreiche Stoffe.

Es ist auffällig, dass die Rangfolge der Feature-Gewichtung der mobilen und sehr mobilen Wirkstoffe der Rangfolge, die sich bei den Metaboliten zeigt, ähnelt, auch wenn diese Ergebnisse aufgrund der z.T. nicht ausreichenden Modellgüten mit Unsicherheiten behaftet sind. Unterschiede zeigen sich zwischen Metaboliten- bzw. mobilen Wirkstoffmodellen und den moderat mobilen Wirkstoffmodellen hinsichtlich der Rangfolge der Feature-Gewichtung. Boden- und hydrogeologische Features gewinnen bei den moderat mobilen Wirkstoffen an Bedeutung während das Feature Ackerland in den Hintergrund gerät. Den flachen Feature-Effekten nach zu urteilen scheint eine große Fläche an Ackerland im Umkreis der Messstelle alleinig nicht auszureichen um die Konzentrationsklasse korrekt vorherzusagen. Es werden vielmehr auch bodenkundliche und hydrogeologische Standortfaktoren herangezogen.

Die Filteroberkante ist nach wie vor ein wichtiges Feature und der Zusammenhang, dass eine oberflächennahe Filteroberkante und eine höhere Wahrscheinlichkeit für Konzentrationsklasse 2 assoziiert ist, besteht weiterhin. Die Makroporenklasse zeigt bei diesen Modellen eine geringere Feature-Gewichtung als der Sandgehalt. Wir gehen davon aus, dass sich – wie für die Metabolitenmodelle diskutiert - vermutlich die Gewichtung des kontinuierlichen Sandgehalts gegenüber den kategorischen Makroporenklassen methodisch bedingt erhöht. Bei den moderat mobilen Wirkstoffen zeigt sich, dass hinsichtlich der Feature-Effekte die Wahrscheinlichkeit höherer Konzentrationen mit dem Potenzial an Makroporenbildung zunimmt, während ein abnehmender Sandgehalt in Zusammenhang mit höheren Konzentrationen steht. Somit ist der Einfluss der Makroporenklassen komplementär zum Sandgehalt zu verstehen. Ob das Modell den Prozess Makroporenfluss erlernt, ist auch bei den moderat mobilen Wirkstoffen wie bei den betrachteten Metaboliten schwach belegt.

Bei den Modellen der erweiterten Feature-Auswahl werden grundwasserchemische Features, vor allen Dingen der Nitratgehalt, als bedeutend eingestuft, was sich mit den Ergebnissen der Metabolitenmodelle deckt. Dies zeigt, dass Nitrat als Zeiger der landwirtschaftlichen Intensität bei Metaboliten und Wirkstoffen zur verbesserten räumlichen Identifikation von potenziell belasteten Messstellen beiträgt. Auf die Besonderheiten von Wirkstoffen und ihre Einordnung in den Kontext der Vulnerabilität wird weiterhin in Kapitel 5.3 eingegangen.

4.3 Diskussion zur Datenverfügbarkeit

In diesem Projekt wurde eine einzigartige Datenlage geschaffen, welche die Auswertung von Wirkstoff- bzw. Metabolitenkonzentrationen an tausenden von Messstellen in Zusammenhang mit Standortfaktoren ermöglicht. Gleichzeitig mussten aufgrund von Einschränkungen bei Datenverfügbarkeit und -qualität einige Entscheidungen darüber getroffen werden, welche Messstellen und Stoffe letztlich in die Analyse einbezogen wurden. Im Folgenden möchten wir einen zusammenfassenden Ausblick auf mögliche, zukünftige Verbesserungen bezüglich der Datenerhebung und -Handhabung geben.

Bei der Verarbeitung der Stoffkonzentrationen war aufgefallen, dass oftmals die LOD (Nachweisgrenze) nicht verfügbar ist. Da sowohl LOQ als auch LOD seitens der Labore bestimmt werden, empfehlen wir für die zukünftige Datenhaltung, diese Informationen für jede Messung

mit aufzunehmen, um eine verbesserte Auswertung linkszensierter Daten gewährleisten zu können.

Zahlreiche aussagekräftige hydrogeologische Standorteigenschaften liegen derzeit nicht bundesweit harmonisiert mit der benötigten räumlichen Abdeckung und Datenqualität vor, um diese in Modelle zu integrieren. Viele Features, die in anderen Vulnerabilitätskonzepten zum Einsatz kommen (z.B. DRASTIC) standen nicht im notwendigen Umfang zur Verfügung. Zu nennen sind die Petrographie und Eigenschaften der vadosen Zone, sowie Informationen zum Grundwasserflurabstand. Oftmals sind Datenerhebung und -haltung in Bezug auf Stoffmessungen und Grundwasserstände nicht aufeinander abgestimmt. Das bedeutet, dass an vielen Messstellen zur Überwachung der Grundwasserqualität keine zeitgleichen Informationen zum Grundwasserflurabstand verfügbar sind. Selbst wenn diese Daten vorhanden sind, sind sie oftmals nicht in einer Datenbank verbunden, sondern müssen zusammengeführt werden. Dieser Aufwand erscheint vermeidbar. Gleichzeitig mussten Informationen zur Grundwasserüberdeckung anhand von Schichtdaten aufgrund des unverhältnismäßigen Harmonisierungsaufwandes (SEP2, SEP3, Länderspezifische Schlüssel) im Projekt unberücksichtigt bleiben. Grundsätzlich liegt hier Potenzial zur Integration solcher Daten in der Zukunft.

Des Weiteren sind die Anstromrichtung an den Messstellen und Grundwassergleichenpläne für eine belastbare Vulnerabilitätsbetrachtung zielführend, liegen aber als bundesweit vereinheitlichtes Datenprodukt nicht vor. Die Harmonisierung dieser Informationen ist daher zukünftig besonders wichtig. Weitere Ansätze, die wichtige Informationen zum Stofftransport und der Verweilzeit erschließen, etwa Schätzungen des Grundwasseralters aus Isotopen oder die Auswertung der Grundwasserganglinien, um die Reaktion des Systems auf Eintrag durch Versickerung zu verstehen, können wertvolle Beiträge leisten. Gleichzeitig unterliegen diese Ansätze wieder den genannten Einschränkungen durch geringe Überlappung zwischen Messstellen, an denen diese Analysen möglich sind, und den Messstellen der Grundwasserqualitäts-Überwachung. Um die Aussagekraft hydrogeologischer Daten zu erhöhen und die intrinsische Vulnerabilität von Standorten fundiert zu bewerten, braucht es eine bundesweite Abstimmung der zugrundeliegenden Datenstrukturen und Methoden.

5 Bewertung der Vulnerabilität

Im Folgenden werden die zuvor vorgestellten und diskutierten Modellergebnisse in den Kontext der Vulnerabilitätsbewertung eingebettet und interpretiert. In diesem Kapitel werden daher zuerst die Grundlagen der Bestimmung der intrinsischen Vulnerabilität behandelt und ein Kartenprodukt zur Bewertung der Schutzfunktion der Grundwasserüberdeckung (auf Grundlage der Methoden nach Hölting et al. (1995) und der LAWA) mit den Stoffkonzentrationen im FARM-Datensatz abgeglichen. Weiterhin wird ein solcher Abgleich auch für prozessorientierte Modelle (Simulationen mit GeoPELMO DE) durchgeführt. Die Vulnerabilitätsbewertung des Grundwassers ist besonders wichtig, wenn Monitoringdaten in der Risikobewertung von PSM-Anwendungen verwendet werden und für Zulassungs- oder Managemententscheidungen in der Regulatorik genutzt werden. Eine Quantifizierung von Vulnerabilität ist aus regulatorischer Sicht notwendig aber komplex in der Umsetzung. Die Ergebnisse dieser Abgleiche sowie deren Implikationen werden im Folgenden gemeinsam mit den Erkenntnissen aus den ML-Modellen diskutiert.

5.1 Verfahren zur Bestimmung der intrinsischen Vulnerabilität

5.1.1 Hintergrund

In hydrogeologischen Systemanalysen werden lithologische Karten und Grundwasserneubildungskarten kombiniert (Albinet et al. 1970; Margat 1968). Weiterhin finden Parametermodelle Anwendung, die u.a. qualitative Bewertungsverfahren nach Punktesystem oder Indexverfahren sein können. Eine dritte Kategorie sind mathematische und statistische Modelle (Magiera 2000). Zu den Analogie-Modellen zählen u.a. der Attenuation Factor (AF; (Rao et al. 1985) und der Leaching Index (LI; (Bacci et al. 1993), welche aus vereinfachten Transportgleichungen abgeleitet wurden. Ein Modell, das die intrinsische Vulnerabilität bewertet, ist der Aquifer Vulnerability Index (AVI; (Stempvoort et al. 1993). Die Gruppe der Punktebewertungs- bzw. Indexverfahren ist am weitesten verbreitet. Diese Methoden nutzen eine Anzahl sorgfältig ausgewählter Parameter. Für jeden Parameter werden Wertebereiche definiert, welche mit einem Rating versehen werden. Zusätzlich wird jedem Parameter eine Wichtung in Form eines Faktors zugewiesen. Die Ergebnisse werden addiert, um die endgültige Punktzahl zu erhalten. Diese Punktzahl liefert ein relatives Maß für die Grundwasservulnerabilität gegenüber Schadstoffeinträgen. Beispiele für Indexverfahren sind das von der EPA entwickelte DRASTIC-Verfahren (Aller et al. 1987), die SINTACS-Methode (Civita et al. 2008) und die in Deutschland verbreitete Hölting-Methode (Hölting et al. 1995). Diese Methoden werden häufig als Grundlage für Weiterentwicklungen genutzt, wie bspw. die Kombination des Hölting-Verfahrens mit der bodenkundlichen Filter- und Pufferfunktion (Wirsing et al. 2015). Allerdings wird in dieser Methodengruppe fast ausschließlich die intrinsische Vulnerabilität eines Aquifers bewertet. Im Kontext der Vulnerabilitätsbewertung gegenüber Nitrateintrag werden für die Verweilzeitermittlung im Bodenbereich und den Deckschichten im Lockergesteinsbereich das Verfahren nach DIN 19732 (Deutsches Institut für Normung 1997) sowie für die Grundwasserdeckschichten in Festgesteinsregionen das Verfahren nach Hölting et al. (1995) empfohlen (LAWA 2018). Eine detailliertere Übersicht und Gegenüberstellung der Verfahren zur Bestimmung der intrinsischen Vulnerabilität findet sich in der Tabelle 9 im Anhang A.8. Je nach Verfahren werden unterschiedliche Eingangsdaten benötigt zur Bewertung der intrinsischen Vulnerabilität. Diese Eingangsdaten werden differenziert nach Bodenparameter, Informationen zur ungesättigten Zone, Aquifer-Eigenschaften sowie zusätzlich nicht näher kategorisierbaren Parametern, die für eine Bewertung benötigt werden.

Für Deutschland liegt ein flächendeckendes Kartenprodukt zum Schutzpotenzial der Grundwasserüberdeckung vor (BGR und SGD 2005). Die Schutzfunktion basiert auf einer Bewertung der ungesättigten Zone und des Bodens hinsichtlich des vertikalen Transports von Schadstoffen in den darunterliegenden, oberen zusammenhängenden, wasserwirtschaftlich nutzbaren Grundwasserleiter. Gemäß der jeweiligen Datenverfügbarkeit wurde die Bewertung der Grundwasserüberdeckung in den Bundesländern entweder anhand der LAWA-Arbeitshilfe (LAWA 2003a) zur Umsetzung der WRRL oder anhand der Methodik nach Hölting et al. (1995) durchgeführt. Die Einteilung der Bewertungsklassen wurde in die Einteilung gemäß der LAWA-Methodik übersetzt, sodass Flächen mit einer „günstigen“, „mittleren“ oder „ungünstigen“ Schutzfunktion klassifiziert wurden (BGR und SGD 2016). Weitere Informationen zu diesen Karten finden sich im Anhang. Der nachfolgend vorgestellte Abgleich mit dem FARM-Datensatz basiert auf diesem Datenprodukt.

5.1.2 Abgleich mit Monitoringdaten

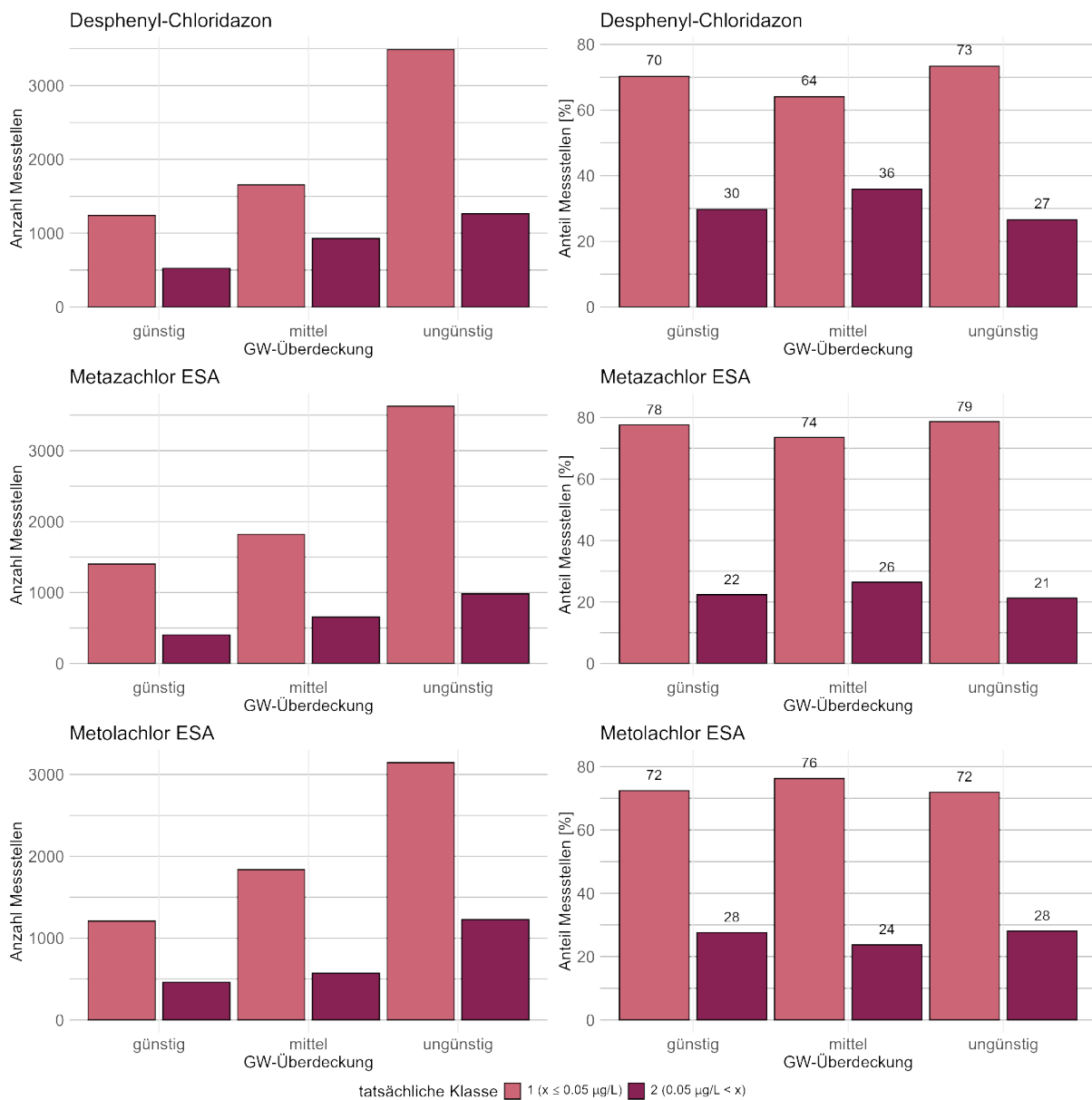
Die Einteilung der Schutzfunktion wurde an den Koordinaten der Grundwassermessstellen extrahiert und mit den Funden von Wirkstoffen und Metaboliten verglichen. Hierbei wurden folgenden Vergleiche mit der Schutzfunktion vorgenommen:

- ▶ *Binäre Konzentrationsklassen für die acht Metaboliten, die in den ML-Modellen evaluiert wurden*
- ▶ *Überschreitungsfrequenzen von 0,01, 0,05 und 0,1 µg L⁻¹ für nach Stoffeigenschaften zusammengefasste Gruppen an Wirkstoffen*

Die Ergebnisse der Metabolite sind in Abbildung 98 für die absolute Anzahl an Messstellen in den jeweiligen Konzentrationsklassen und pro Schutzfunktionsklasse pro Stoff dargestellt (linke Seite der Graphik). Auf der rechten Seite sind jeweils die relativen Anteile der Messstelle pro Schutzfunktionsklasse dargestellt. Obgleich eine größere Anzahl an Messstellen den ungünstigen (vulnerablen) Standorten zuzuordnen ist, sind die relativen Unterschiede vernachlässigbar. Hier werden beispielhaft Desphenyl-Chloridazon, Metazachlor ESA und Metolachlor ESA gezeigt, diese Beobachtung konnte jedoch für alle acht untersuchten Metaboliten gemacht werden. Der Abgleich der binären Konzentrationsklassen für die acht mittels der ML-Modelle evaluierten Metaboliten mit den drei Klassen der Schutzfunktion der Grundwasserüberdeckung zeigt keinen Zusammenhang. Die Konzentrationsklasse 2 wird an ungünstigen (vulnerablen) Standorten anteilig nicht häufiger gefunden als an Standorten mit günstiger Schutzfunktion.

In Abbildung 99 wird der Abgleich für alle Wirkstoffe zusammengefasst und für Überschreitungen über 0,1 µg L⁻¹ gezeigt. Hierbei zeigt sich ein prozentual größerer Anteil an Messstellen mit Funden an den ungünstigen Standorten als an den günstigen, die als mittel geschützt eingestuft sind. Alle drei nach Mobilität aufgeteilten Wirkstoffgruppen zeigen ebenso diese Abstufung mit dem höchsten Anteil an Messstellen mit Funden an ungünstigen Standorten (Abbildung 100, Abbildung 101, Abbildung 102). Der Unterschied zwischen den günstigen und ungünstigen Standorten ist bei den wenig mobilen Wirkstoffen am stärksten ausgeprägt. Auch die Einzelstoffabgleiche bestätigen das Bild, dass bei den Wirkstoffen ein höherer Prozentsatz an Messstellen mit Stoffeinträgen an ungünstigen Standorten liegt. Dies wurde u.a. bei Terbutylazin und Glyphosat beobachtet (hier nicht gezeigt).

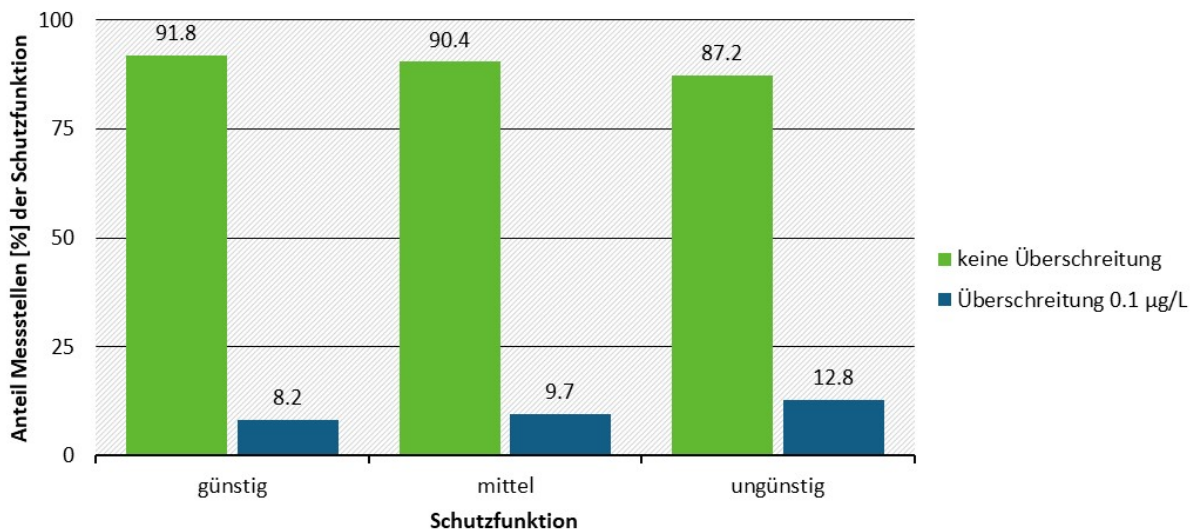
Abbildung 98: Anzahl der Messstellen (links) und prozentuale Anteil an Messstellen (rechts) mit und ohne Überschreitungen des $0,05 \mu\text{g L}^{-1}$ Konzentrationswertes durch die Wirkstoffe Desphenyl-Chloridazon, Metazachlor ESA und Metolachlor ESA in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.



Quelle: eigene Darstellung BGR.

Abbildung 99: Der prozentuale Anteil an Messstellen mit und ohne Überschreitungen des $0,1 \mu\text{g L}^{-1}$ Grenzwertes durch Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung wird dargestellt.

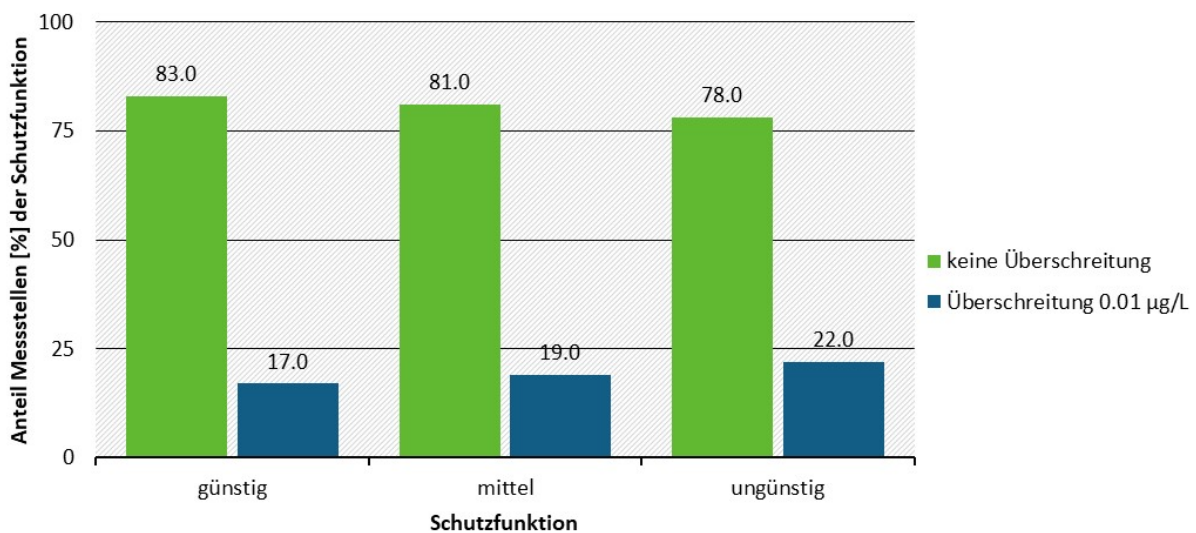
stoffübergreifende Wirkstoffe



Quelle: eigene Darstellung BGR.

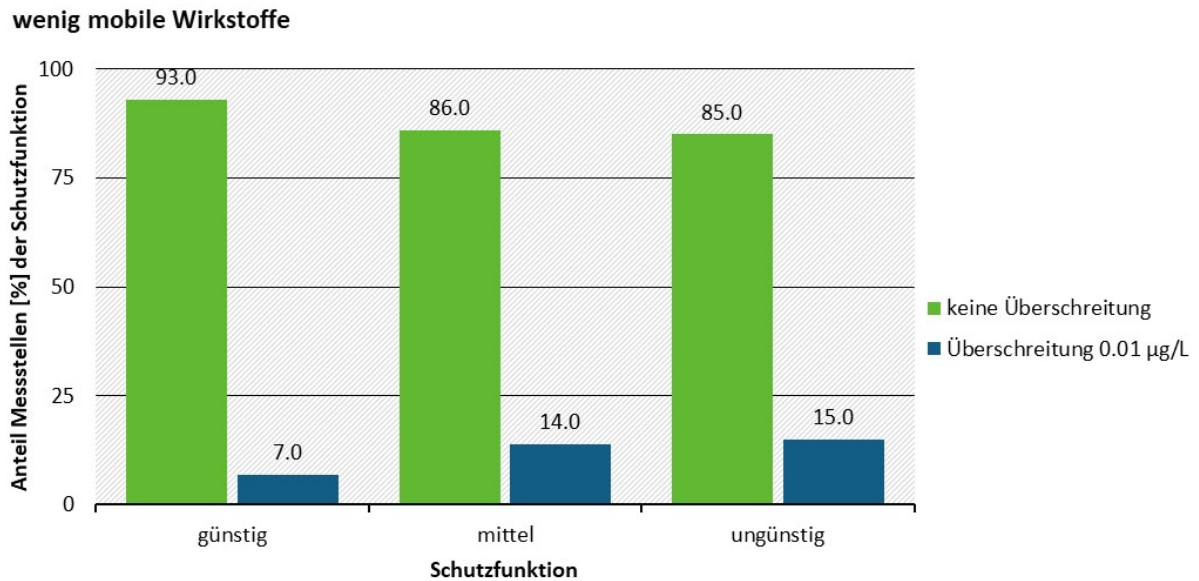
Abbildung 100: Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des $0,01 \mu\text{g L}^{-1}$ Konzentrationswertes durch sehr mobile und mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.

sehr mobile und mobile Wirkstoffe



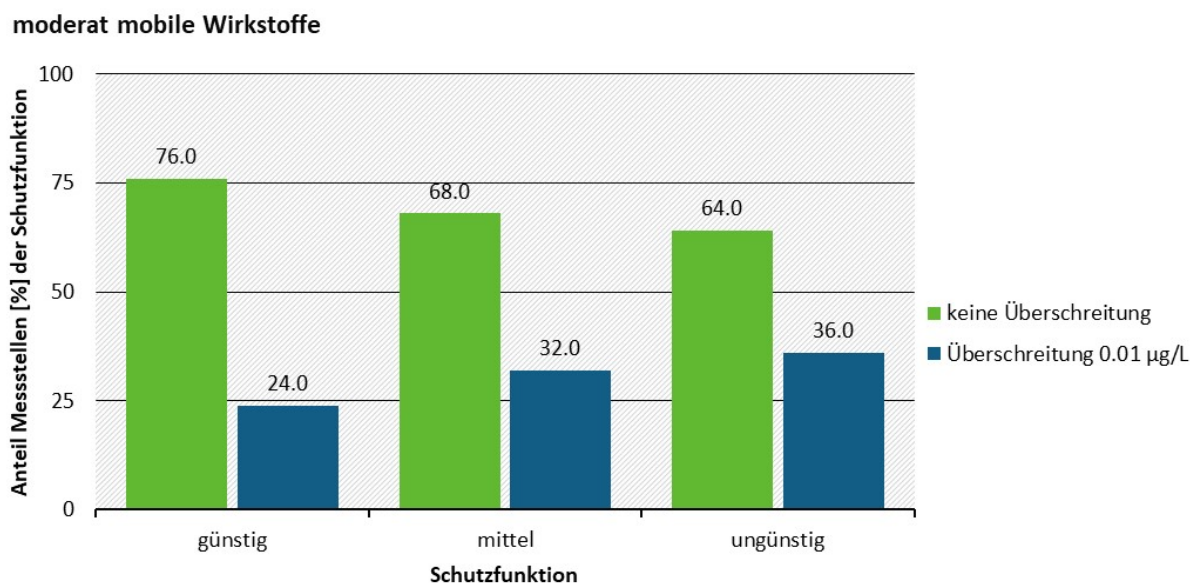
Quelle: eigene Darstellung BGR.

Abbildung 101: Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des $0,01 \mu\text{g L}^{-1}$ Konzentrationswertes durch wenig mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.



Quelle: eigene Darstellung BGR.

Abbildung 102: Balkendiagramm des prozentualen Anteils an Messstellen mit und ohne Überschreitungen des $0,01 \mu\text{g L}^{-1}$ Konzentrationswertes durch moderat mobile Wirkstoffe in den drei Kategorien der Schutzfunktion der Grundwasserüberdeckung.



Quelle: eigene Darstellung BGR.

5.2 GEOPELMO DE – Abgleich mit Daten und Vorhersagen

5.2.1 Hintergrund

Die Verwendung unterschiedlicher Ansätze für eine Vulnerabilitätsbewertung von Monitoringmessstellen und -standorten im Kontext Pflanzenschutz werden auf EU-Ebene intensiv diskutiert (Gimsing et al. 2019; Tiktak et al. 2020). Dabei werden räumliche Modellansätze auf Basis der FOCUS-Modelle PEARL, PELMO und MACRO für eine Vulnerabilitätsbewertung in Betracht gezogen, die das Abbau- und Adsorptionsverhalten von PSM im Boden in Abhängigkeit von Anwendungsmenge und -zeitraum prozessorientiert berücksichtigen und eine zu erwartende Versickerungsneigung berechnen. Diese Modelle gehören zur Gruppe der *spatial distributed leaching models* (SDLM). Der Vorteil dieser vorgeschlagenen Ansätze ist, dass sie stoffspezifische Eigenschaften sensitiv berücksichtigen und damit der Bewertung einer spezifischen Standortvulnerabilität gegenüber PSM generell zuträglich sind. Darin unterscheiden sie sich von einfachen Indexansätzen, die vorwiegend eine intrinsische Vulnerabilität abbilden. Allerdings wurde wiederholt Kritik geäußert und exemplarisch anhand von standortspezifischen Monitoringergebnissen diskutiert, dass der Stofftransport in den regulatorisch angewendeten Modellen PEARL und PELMO überwiegend am Matrixfluss in gut durchlässigen Böden kalibriert ist. Dies kann dazu führen, dass die räumliche Extrapolation der Modellannahmen auf bindige Böden mit Makroporenfluss stark fehlerhaft sein kann (Rosenbom et al. 2015). Für Deutschland liegen Modellerweiterungen für PELMO vor, die Oberflächenabfluss, Makroporenfluss und Drainage im Boden berücksichtigen und für flächenhafte Validierungen zur Verfügung stehen. Alle diese für Vulnerabilitätsanalysen in Betracht gezogenen PSM-Modelle bilden jedoch überwiegend die Prozesse in der ungesättigten Bodenzone ab. Sie berücksichtigen kaum den Grundwasserflurabstand bzw. den gesamten Aufbau und die gesamte Mächtigkeit der ungesättigten Zone und die damit zusammenhängenden Versickerungs- bzw. Verweilzeiten. Hydrogeologische Gegebenheiten der gesättigten Grundwasserzone werden gänzlich unberücksichtigt gelassen. Darin liegt eine starke Begrenzung dieser Modellansätze begründet, wenn es um die Auswahl von vulnerablen Standorten für zielgerichtete Monitoringstudien und die Interpretation von Monitoringergebnissen geht. Ein weiteres Defizit ist, dass die existierenden räumlichen Modellansätze bisher kaum anhand größerer Monitoringdatensätze validiert wurden. Damit fehlt ein Abgleich, ob die empfohlene Risikobewertung auf Basis der FOCUS-Modelle die Realität hinreichend gut abbilden und ob bzw. in welchen Bewertungszusammenhängen sie letztlich auch national genutzt werden können. Daher wurde untersucht, ob einzelstoffspezifische Modellierungen der Versickerungsneigung das Vorkommen von höheren Konzentrationen im Grundwasser im Sinne einer Vulnerabilitätsvorhersage plausibel erklären können, indem diese mit dem im FARM-Projekt genutzten Datensatz verglichen werden.

Die Spezifikationen der Modellläufe finden sich in UBA-Berichten (Klein et al. 2019; Klein et al. 2025).

Für eine Auswahl an Wirkstoffen und Metaboliten liegen Simulationsergebnisse mit GeoPELMO-DE für jeweils zwei Modellkonfigurationen vor:

- ▶ *GeoPELMO DE FOCUS L: Leaching, d.h. Versickerung durch Matrixfluss (chromatographisches Fließen entsprechend (FOCUS 2009/2014)*
- ▶ *GeoPELMO DE FOCUS L+R+M+D: Leaching + Runoff (Oberflächenabfluss) + Makroporenfluss + Drainage*

Die verwendeten simulierten Konzentrationen in der Fläche stellen jeweils das 80. zeitliche Perzentil der mittleren jährlichen Sickerwasserkonzentrationen in 1 m Bodentiefe und zeitlich aus 20 Modelljahren (1991-2010) dar. Es wurde der Mittelwert der Pixel berechnet, die sich im 1 km-Radius um die Messstelle befinden. Anhand dieser messstellenbezogenen Werte wurden wiederum räumliche Perzentile berechnet. Somit erfolgte eine Einteilung der 80. zeitlichen Perzentile in 4 räumliche Perzentilklassen (0-25., 25.-50., 50.-75., 75.-100.). Für vier Wirkstoffe und sieben Metaboliten, für die eine verwertbare Anzahl von Messwerten in der FARM-Datenbank bereitstand, wurden folgende Abgleiche durchgeführt:

Metaboliten:

Abgleich zwischen den relativen Anteilen der Messstellen in Konzentrationsklassen 1 ($< 0,05 \mu\text{g L}^{-1}$) und 2 ($> 0,05 \mu\text{g L}^{-1}$) für sieben Metaboliten aus den ML-Modellen: Desphenyl-Chloridazon, Methyl-desphenyl-Chloridazon, Metazachlor ESA, Metazachlorsäure, Metolachlor ESA, Metolachlor OXA und NOA 413173.

Wirkstoffe:

Abgleich zwischen den relativen Anteilen der Messstellen mit und ohne Konzentrationen $> 0,01 \mu\text{g L}^{-1}$ bzw. $> 0,1 \mu\text{g L}^{-1}$ für Bentazon bzw. $> 0,01 \mu\text{g L}^{-1}$ für Metazachlor, Terbutylazin und Metolachlor.

5.2.2 Abgleich der Simulationen mit Monitoringdaten

Es befinden sich Messstellen mit Überschreitungen der ausgewählten Konzentrationen in allen Perzentilklassen, unabhängig von der Wahl der Überschreitungskonzentration oder des Wirkstoffs.

Wider den Erwartungen finden sich für den Wirkstoff Bentazon relativ gesehen etwas mehr Messstellen (4,6%) mit Überschreitungen über $0,1 \mu\text{g L}^{-1}$ in der untersten räumlichen Modellperzentilkategorie (0-25.) des GeoPELMO Modells mit Versickerung (Abbildung 103). Die Anteile an den weiteren Perzentilklassen nehmen ab. Bei der Modellversion mit zusätzlichen Prozessen findet sich ebenfalls ein höherer Prozentsatz an Messstellen in der untersten räumlichen Modellperzentilkategorie (3,3%), der Unterschied zwischen den Perzentilklassen ist geringer (Abbildung 104). Der Abgleich hinsichtlich der Messstellen mit Funden größer $0,01 \mu\text{g L}^{-1}$ zeigt ein vergleichbares Bild, da die niedrigste Perzentilkategorie den prozentual größten Anteil an Messstellen mit dieser Konzentrationsklasse zeigt (Abbildung 105, Abbildung 106). Mit aufsteigender Perzentilkategorie nimmt der Anteil der Messstellen mit entsprechenden Funden ab.

Der Abgleich zwischen Messstellen mit Überschreitungen größer $0,01 \mu\text{g L}^{-1}$ und den Perzentilklassen zeigt bei den Wirkstoffen Metazachlor (Abbildung 107, Abbildung 108) und Metolachlor (Abbildung 109, Abbildung 110) relativ gleichverteilte Anteile an Messstellen.

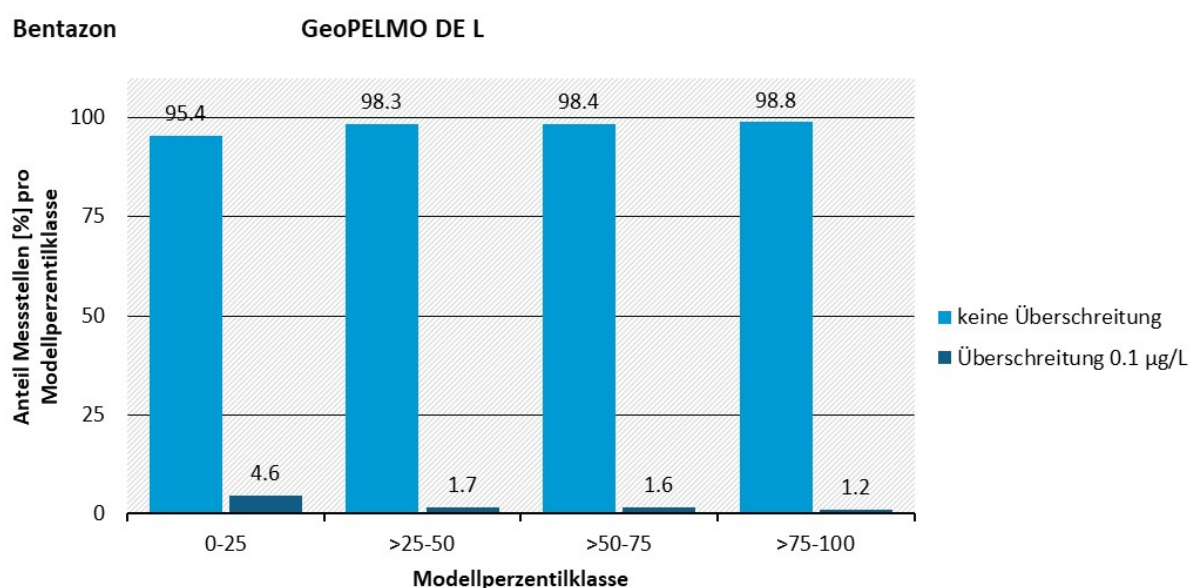
Beim Wirkstoff Terbutylazin (Abbildung 111, Abbildung 112) zeigt sich beim Modell mit ausschließlich Versickerung (FOCUS L) keine eindeutige Tendenz, während sich beim Modell mit den zusätzlichen Prozessen (FOCUS L+R+M+D) eine leichte Tendenz in Richtung höhere Anteile an Messstellen in der höchsten Perzentilkategorie (8,7% im Vergleich zu 4,9% in der niedrigsten Perzentilkategorie) zeigt.

Beim Metaboliten Metazachlor ESA zeigt der Abgleich, dass das Modell mit einfacher Versickerung (FOCUS L) einen höheren Anteil an Messstellen in Konzentrationsklasse 2 ($> 0,05 \mu\text{g L}^{-1}$) in der höchsten Perzentilkategorie aufweist (Abbildung 113). Beim Modell mit den zusätzlichen Prozessen (FOCUS L+R+M+D) zeigt sich kein klarer Zusammenhang (Abbildung 114). Bei Metazachlorsäure zeigen beide Modellkonfigurationen einen leicht höheren Anteil an Messstellen in Konzentrationsklasse 2 in der höchsten Perzentilkategorie (Abbildung 115,

Abbildung 116). Bei Metolachlor ESA (Abbildung 117, Abbildung 118) zeigen beide Modellkonfigurationen einen ähnlich hohen Anteil an Messstellen mit entsprechenden Funden in den ersten drei Perzentilklassen und einen geringeren Anteil in der vierten Perzentilkategorie. Auch bei Metolachlor OXA (Abbildung 119, Abbildung 120) findet sich der geringste Anteil an Messstellen mit Konzentrationen größer $0,05 \mu\text{g L}^{-1}$ in der höchsten räumlichen Perzentilkategorie. Am höchsten ist der Anteil in der zweiten Perzentilkategorie obgleich die Unterschiede nicht sehr ausgeprägt sind. Dies ist für beide Modellkonfigurationen der Fall. Es ist insgesamt herauszustellen, dass der Abgleich mit den Metaboliten keine oder eine leichte Übereinstimmung zeigt, eine starke Gegenläufigkeit allerdings bei den Wirkstoffen (Bentazon) zu beobachten war.

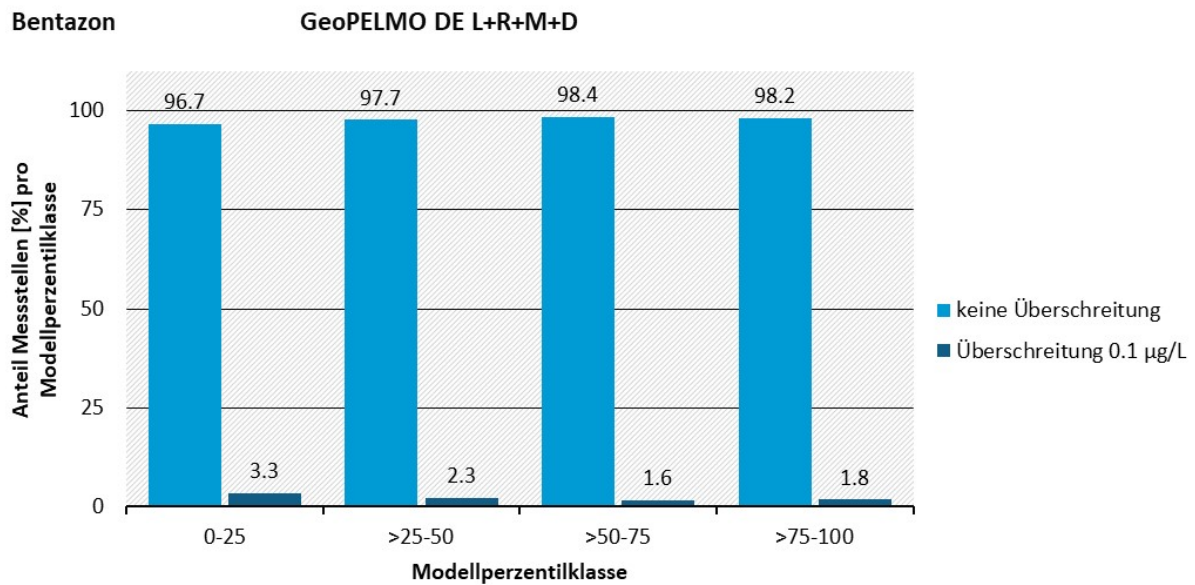
Abschließend lässt sich festhalten, dass eine Diskrepanz zwischen der Vorhersage der SDLM-Modelle für die räumlich-zeitlich aggregierte Perzentilkategorie 0-25 und dem Messstellenabgleich für eine Konzentrationsüberschreitung vorhanden ist. Die Ursachen dafür sind spekulativ und die mechanistischen Zusammenhänge an dieser Stelle nicht zu klären. Sie heben aber den Validierungsbedarf für die Methode SDLM hervor, warum die Wahrscheinlichkeit für eine Konzentrationsüberschreitung als gering oder nicht gegeben für einen Standort klassifiziert wird, es aber dennoch zu einem Eintrag eines PSM-Wirkstoffes oder Metaboliten in das Grundwasser kommt. Vor dem Hintergrund der räumlichen Eingangsdaten, die für die SDLM benötigt werden, ist der räumliche Maßstab ein entscheidendes Kriterium. Eine geeignete Auflösung sollte sich an dem Zielgebiet orientieren (bspw. EU-weit, national oder regional) und die geeignete, verfügbare Datengrundlage berücksichtigen (Tiktak et al. 2020). Auch wenn die Verfügbarkeit von Geodaten zunimmt, sind sie doch häufig eine limitierende Größe für eine kohärente Bestimmung der Grundwasser-Vulnerabilität (Gimsing et al. 2019). Eine Erhöhung der Maßstabszahl von 1:250.000 – als Grundlage im vorstehenden Projekt, fußend auf den hydrologischen und bodenkundlichen Eingangsdaten für das maschinelle Lernen – hin zu 1:1.000.000, wie es EU-weit diskutiert wird (Tiktak et al. 2025), ist für die Vorhersagegüte der Vulnerabilität anhand von SDLMs als kritisch zu bewerten und sollte entsprechend validiert werden.

Abbildung 103: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,1 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L-Modell entnommen wurden, für den Wirkstoff Bentazon.



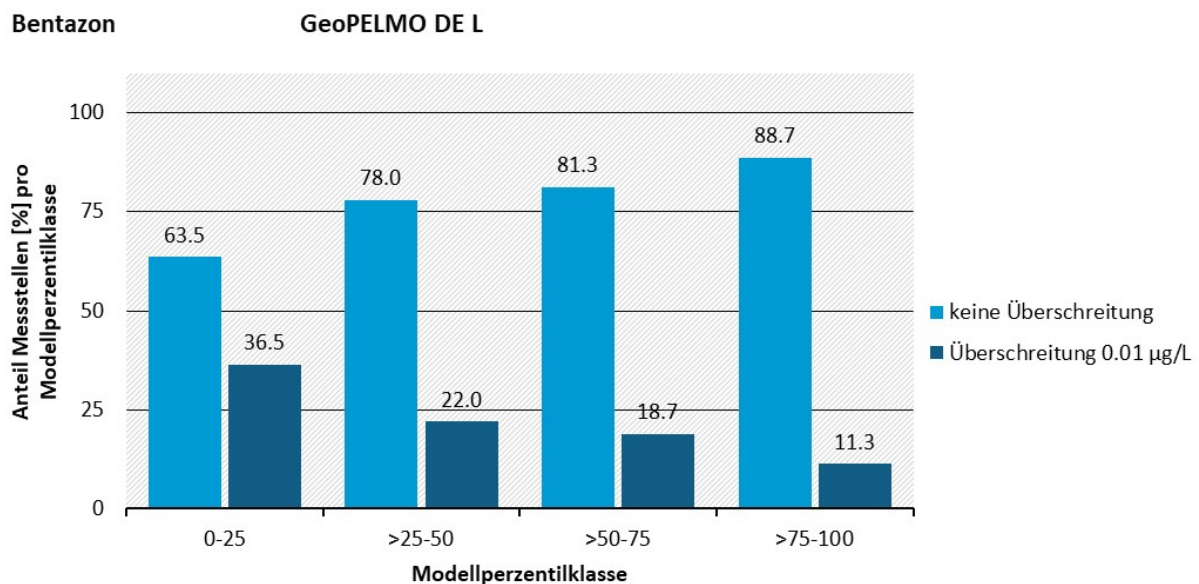
Quelle: eigene Darstellung BGR.

Abbildung 104: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,1 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Bentazon.



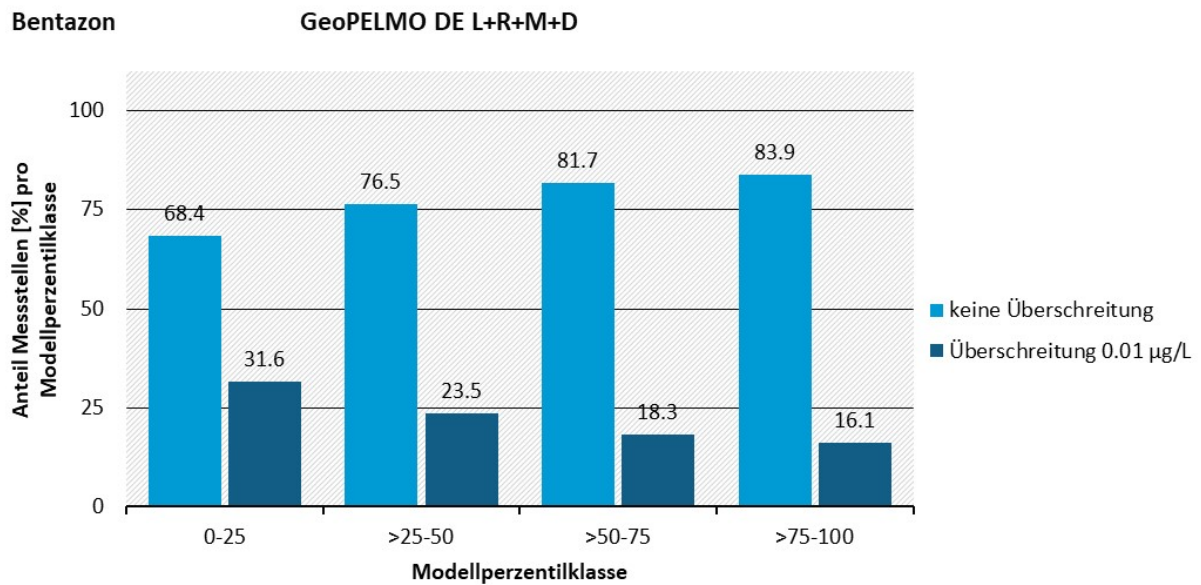
Quelle: eigene Darstellung BGR.

Abbildung 105: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Bentazon.



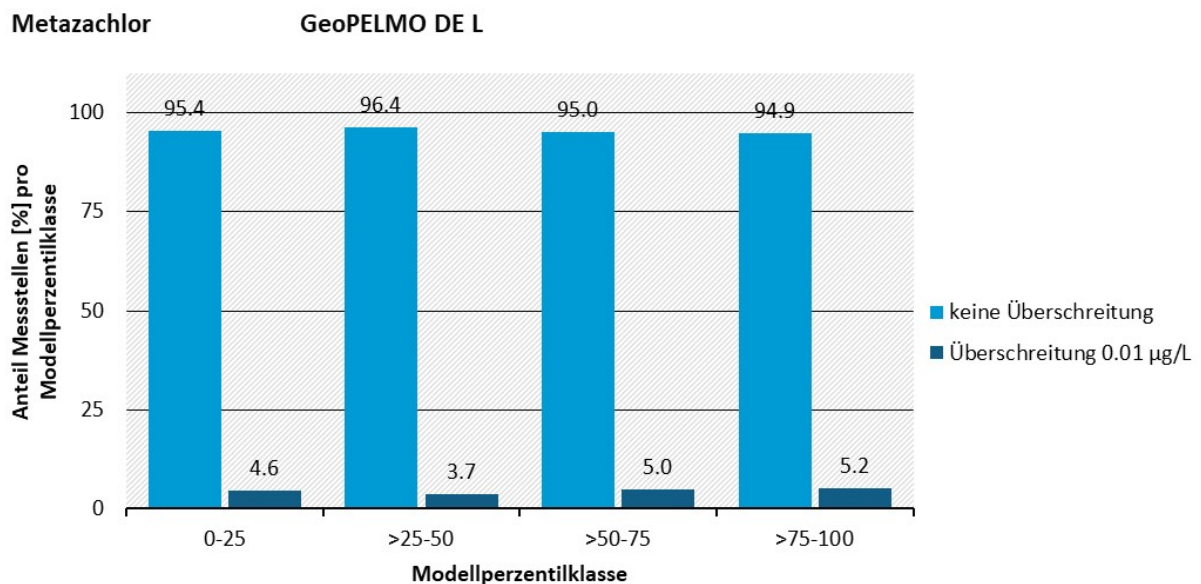
Quelle: eigene Darstellung BGR.

Abbildung 106: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Bentazon.



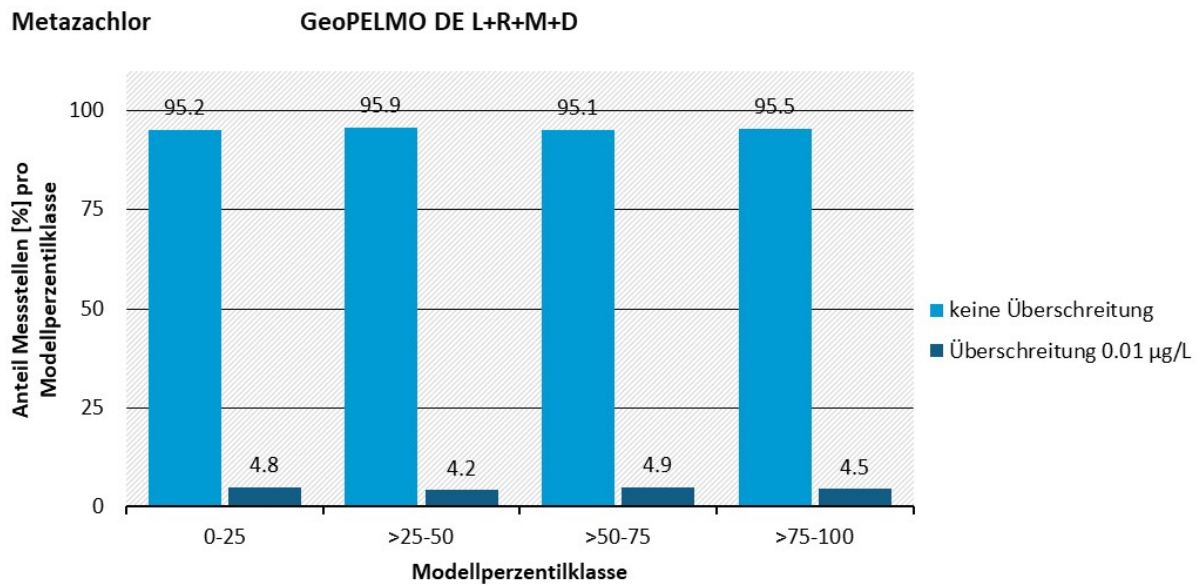
Quelle: eigene Darstellung BGR.

Abbildung 107: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Metazachlor.



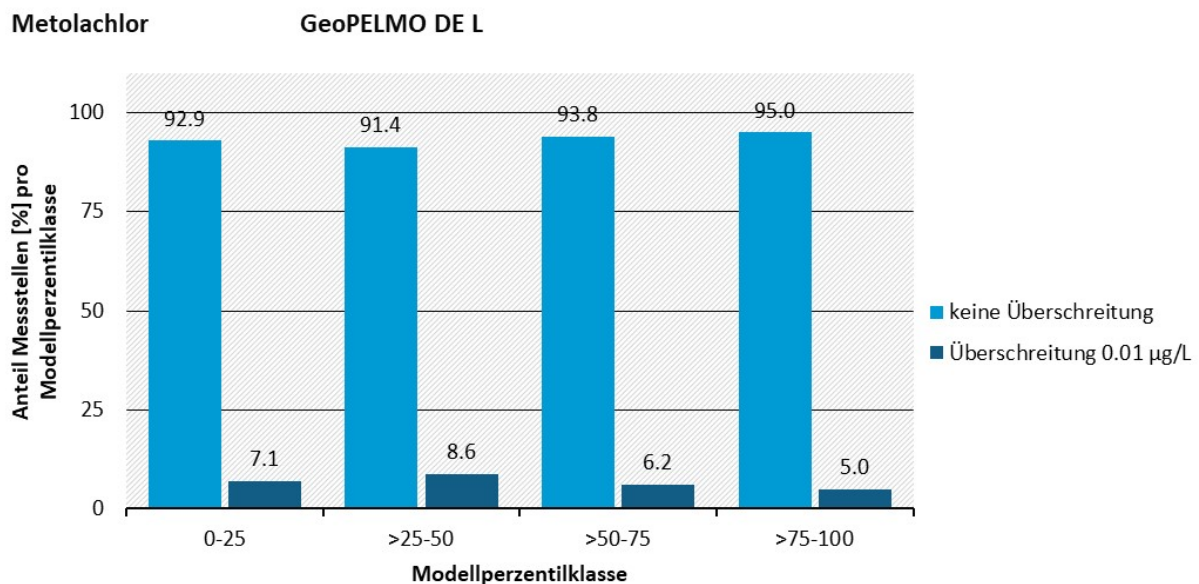
Quelle: eigene Darstellung BGR.

Abbildung 108: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Metazachlor.



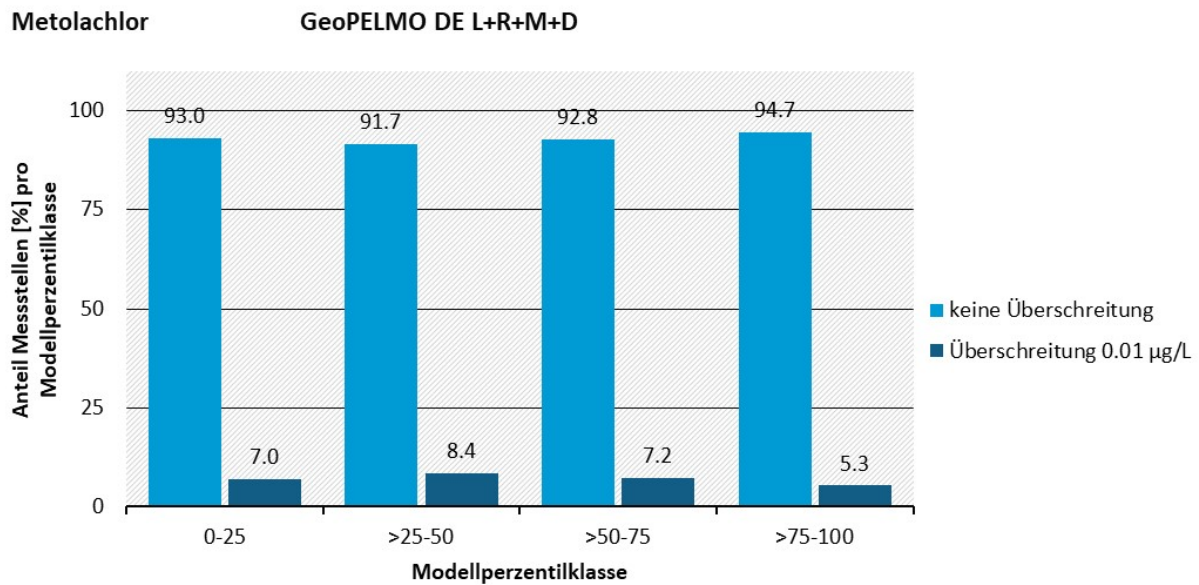
Quelle: eigene Darstellung BGR.

Abbildung 109: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Metolachlor.



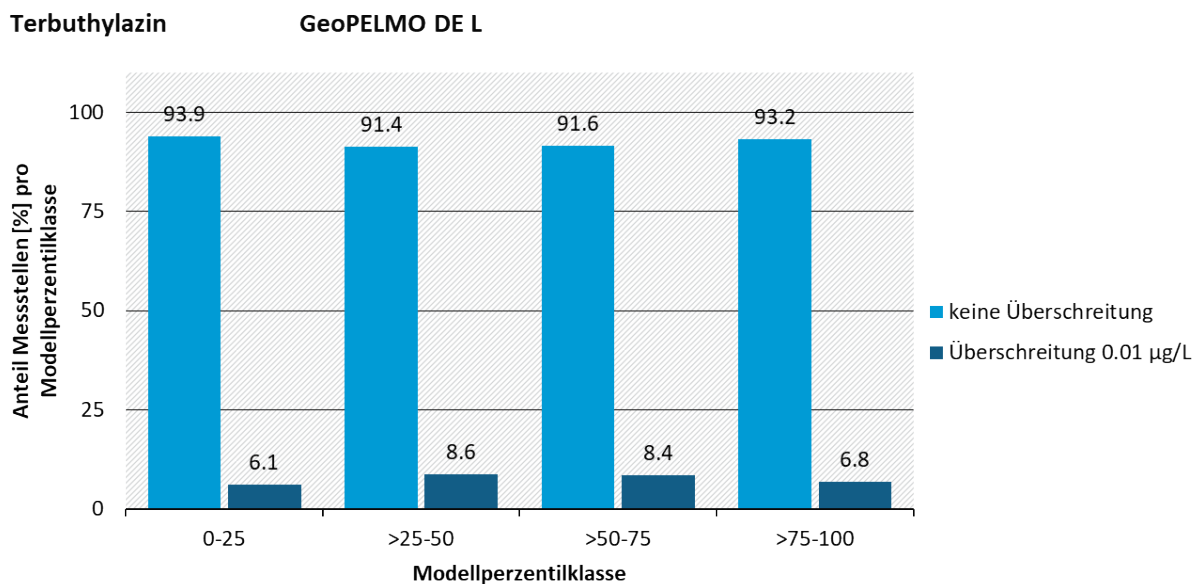
Quelle: eigene Darstellung BGR.

Abbildung 110: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Metolachlor.



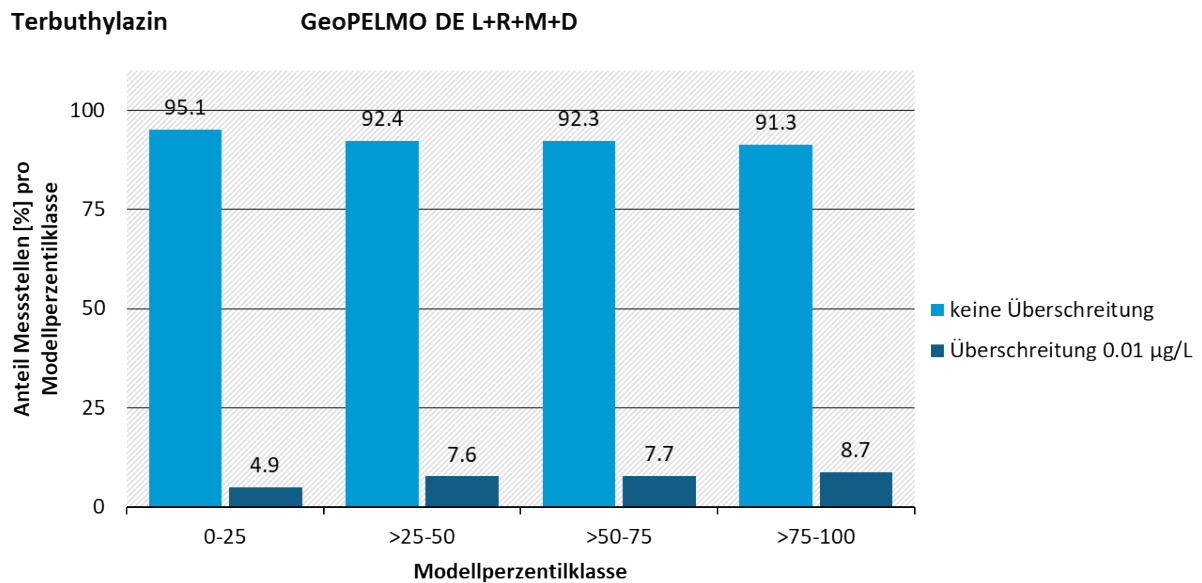
Quelle: eigene Darstellung BGR.

Abbildung 111: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell entnommen wurden, für den Wirkstoff Terbutylazin.



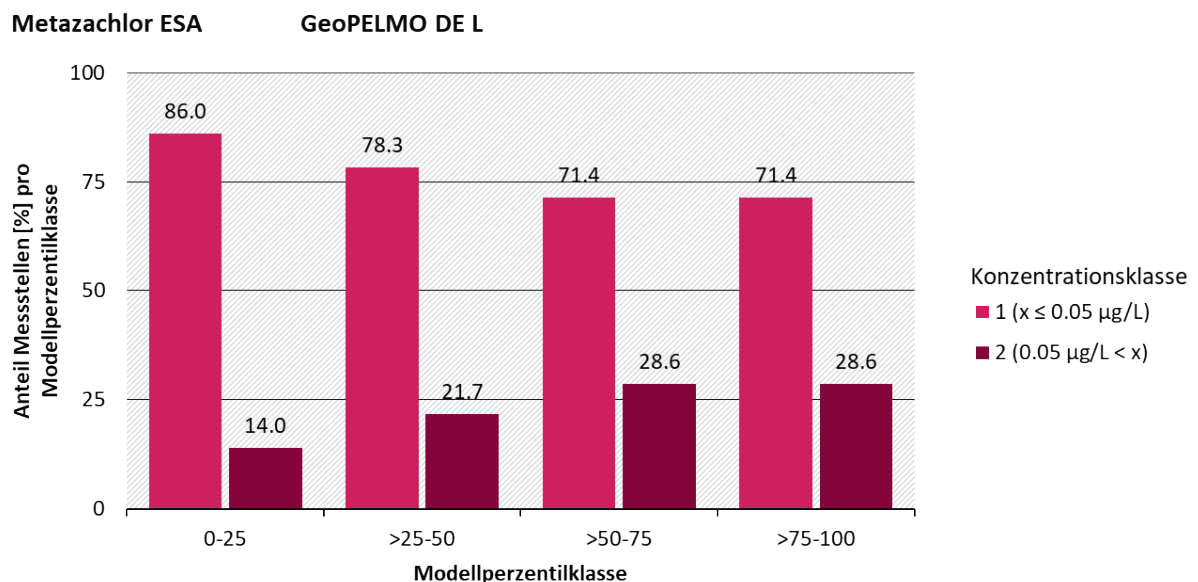
Quelle: eigene Darstellung BGR.

Abbildung 112: Abgleich der relativen Anteile [%] jeweils der Messstellen mit und ohne Überschreitungen von $0,01 \mu\text{g L}^{-1}$ in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen entnommen wurden, für den Wirkstoff Terbutylazin.



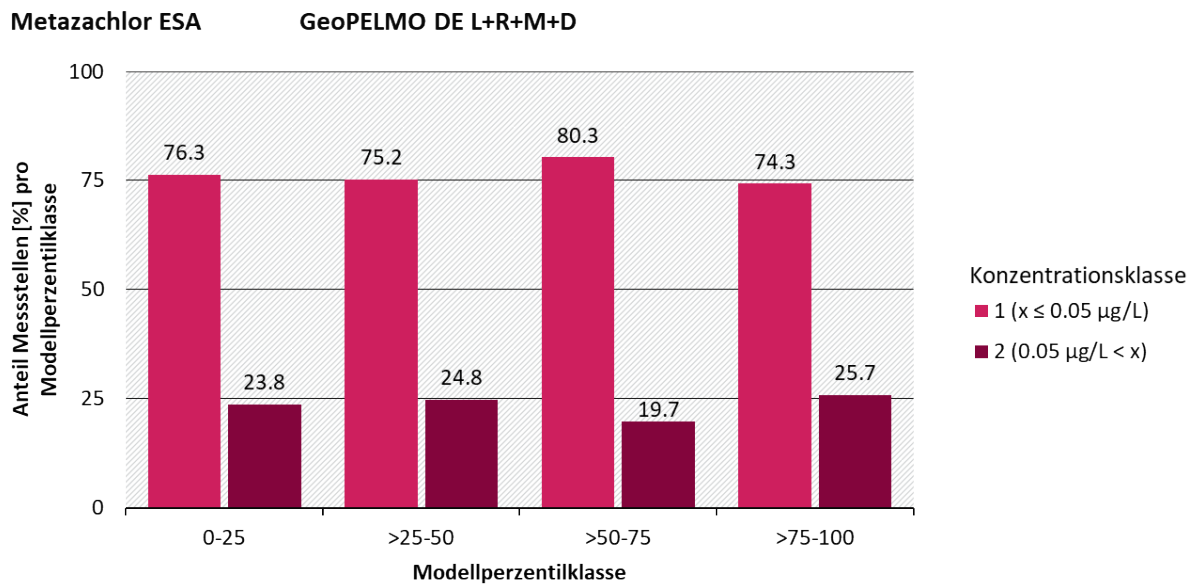
Quelle: eigene Darstellung BGR.

Abbildung 113: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metazachlor ESA.



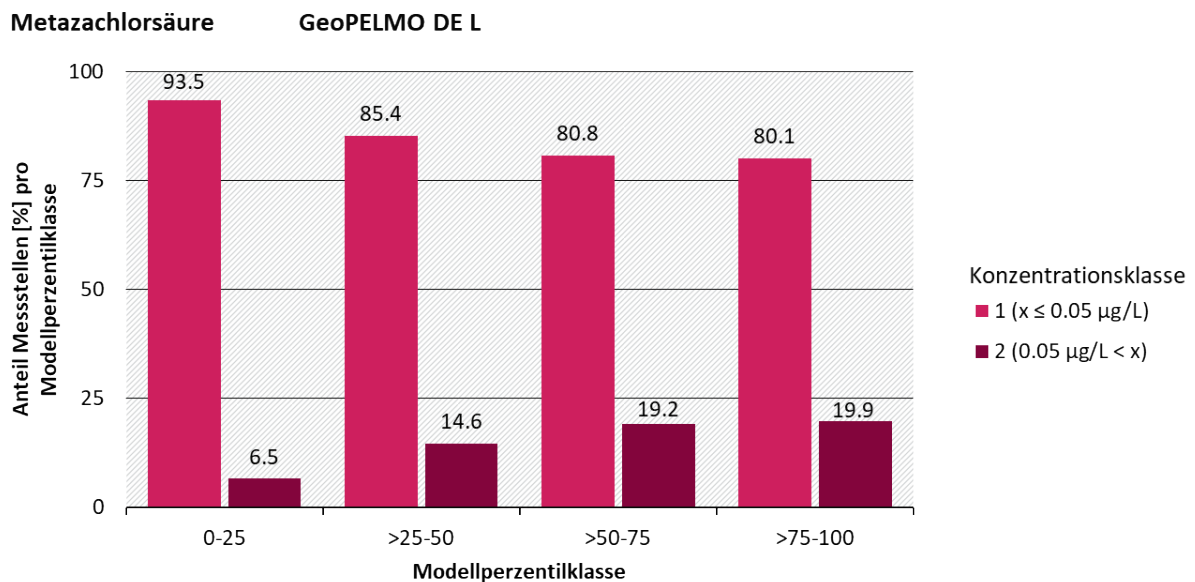
Quelle: eigene Darstellung BGR.

Abbildung 114: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metazachlor ESA.



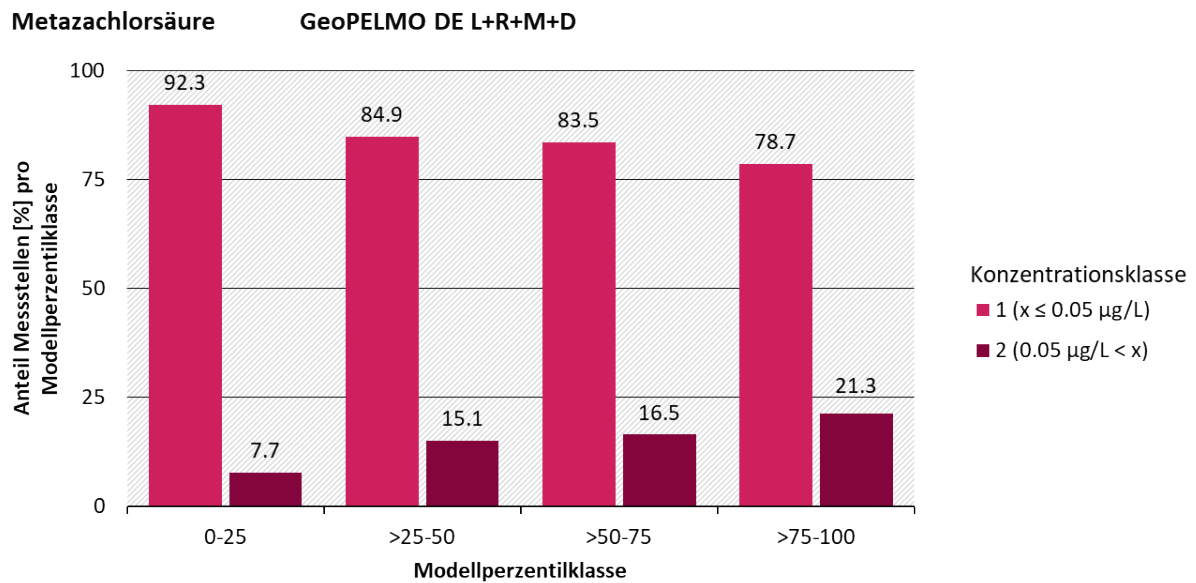
Quelle: eigene Darstellung BGR.

Abbildung 115: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metazachlorsäure.



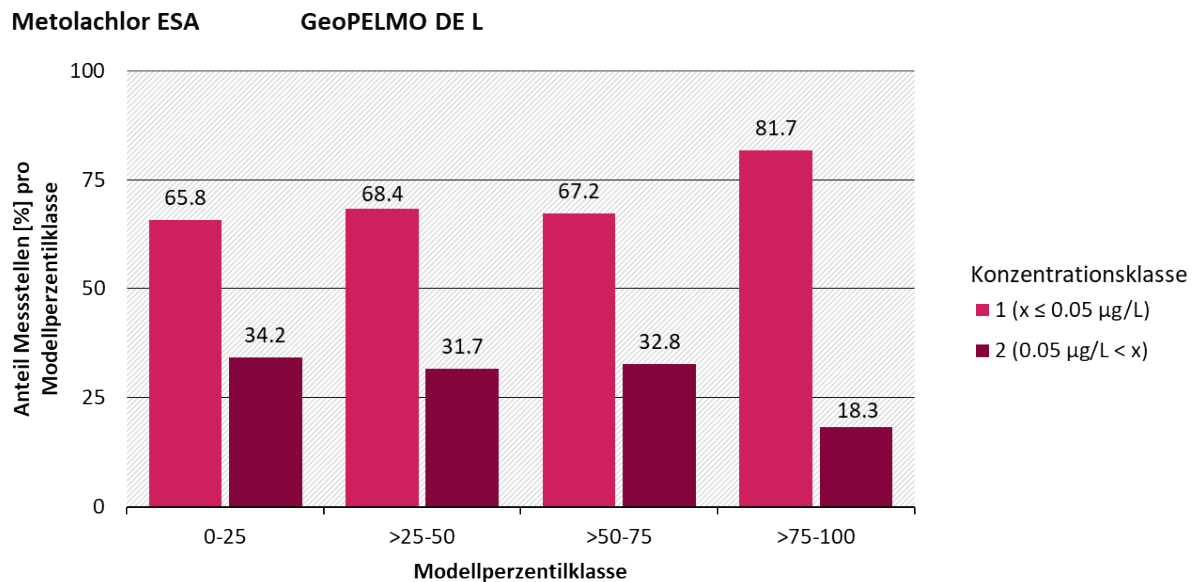
Quelle: eigene Darstellung BGR.

Abbildung 116: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metazachlorsäure.



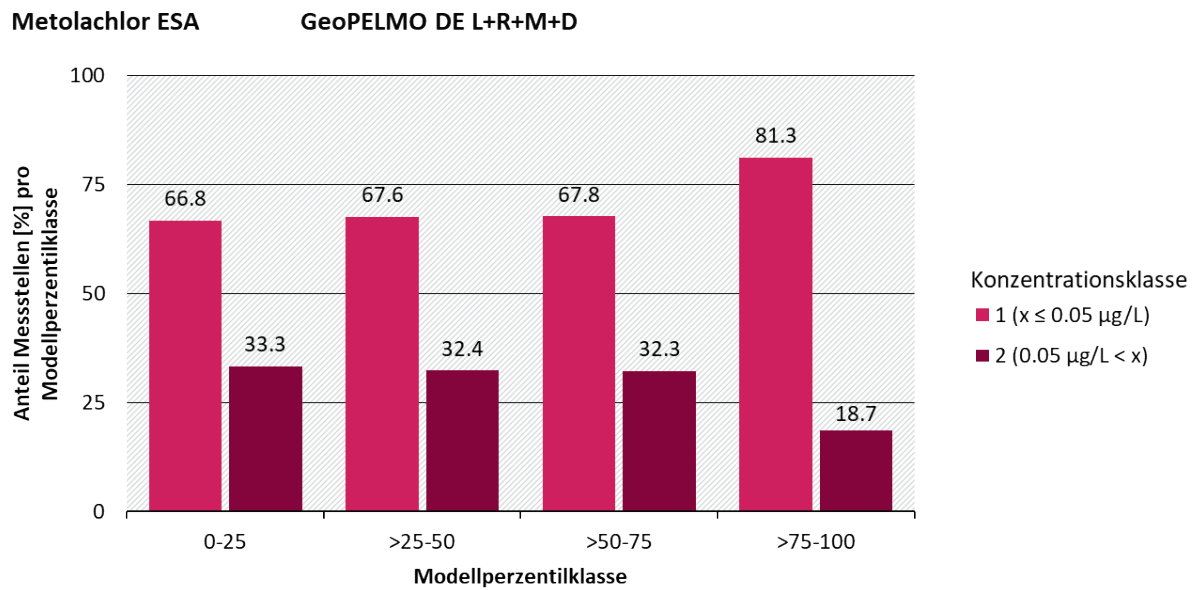
Quelle: eigene Darstellung BGR.

Abbildung 117: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metolachlor ESA.



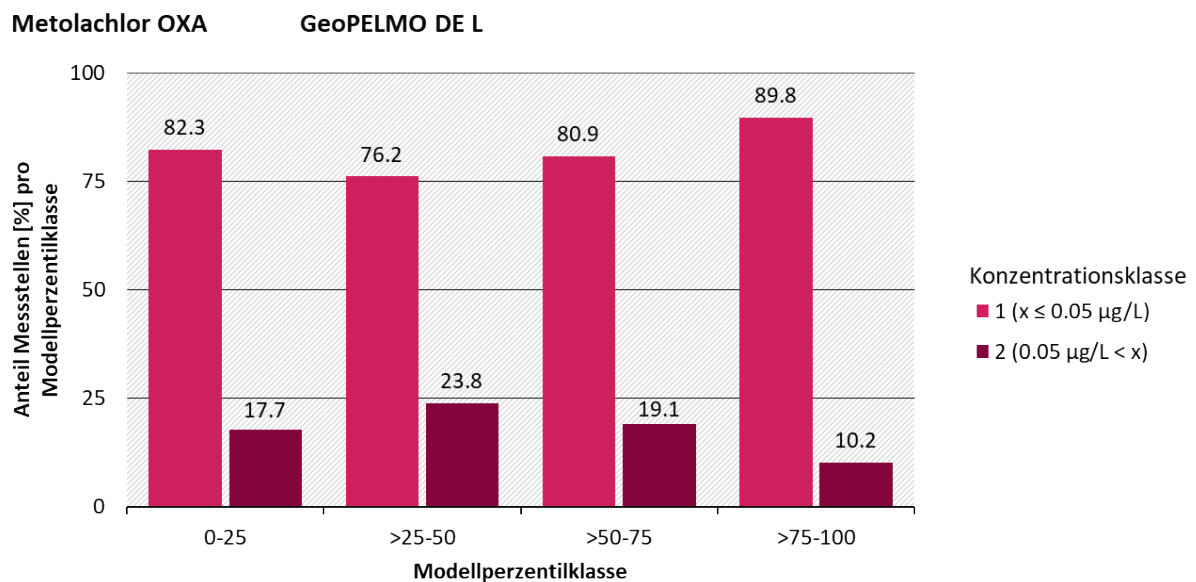
Quelle: eigene Darstellung BGR.

Abbildung 118: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metolachlor ESA.



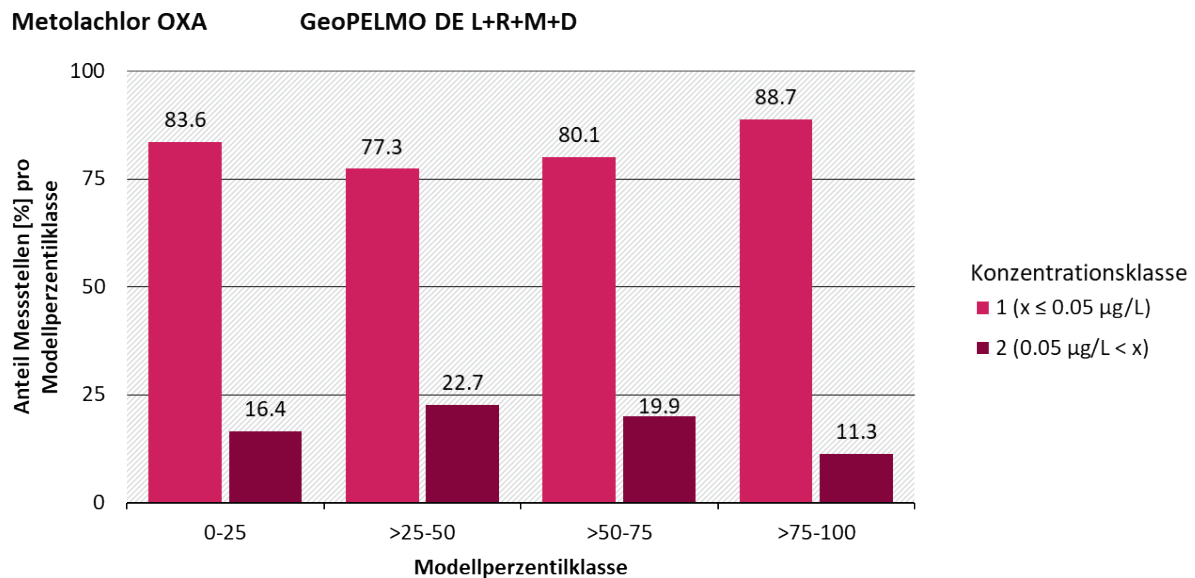
Quelle: eigene Darstellung BGR.

Abbildung 119: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE L Modell, für den Metaboliten Metolachlor OXA.



Quelle: eigene Darstellung BGR.

Abbildung 120: Abgleich der relativen Anteile [%] jeweils der Messstellen in den zwei Konzentrationsklassen in den vier räumlichen Modellperzentilklassen, die aus dem GeoPELMO DE Modell mit zusätzlichen Prozessen, für den Metaboliten Metolachlor OXA.



Quelle: eigene Darstellung BGR.

5.3 Diskussion und Schlussfolgerungen zur Anwendung und Weiterentwicklung von Grundwasser-Vulnerabilitätsindizes

Im Folgenden werden die zuvor gezeigten Abgleiche zwischen den Monitoringdaten und zwei Vulnerabilitätskonzepten sowie die Modellergebnisse einander gegenübergestellt und abschließend im Kontext der Anwendung und Weiterentwicklung von Grundwasser-Vulnerabilitätsindizes diskutiert.

Dabei gehen wir auf die grundsätzliche Frage ein, die zu Beginn des Projektes gestellt wurde:

- *Welche Faktoren bestimmen maßgeblich die Vulnerabilität des Grundwassers gegenüber PSM-Einträgen?*

Die ML-Modelle waren in der Lage relevante Faktoren & räumliche Zusammenhänge aufzudecken, insbesondere der Einfluss der Landwirtschaft und der Filtertiefe wurden als steuernde Größen identifiziert. Für acht mobile und sehr mobile Metaboliten von drei Wirkstoffen stehen oberflächennahe Filter und größere Flächen der zugehörigen Kulturen im Umkreis der Messstellen im Zusammenhang mit höheren Konzentrationen im Grundwasser. Weitere Variablen wie Bodeneigenschaften (Sandgehalt, Makroporen) und die Grundwasserneubildung zeigten je nach Stoff z.T. unterschiedliche Effekte auf die Stoffkonzentrationen. Hydrogeologische Variablen spielten in den Metabolitenmodellen keine Rolle. Diese Modelle liefern kein reines Prozessverständnis zum Versickerungsverhalten und es ist davon auszugehen, dass die Modelle sowohl Prozessverständnis als auch räumliche Korrelationen lernen, die der Verteilung der Anbauflächen innerhalb Deutschlands geschuldet sind. Merkmale wie Bodenparameter, Höhenlage, Niederschlag und Nitratkonzentrationen weisen möglicherweise eher auf eine Korrelation mit Gebieten hin, in denen eine bestimmte Kultur verstärkt angebaut wird, sowie mit landwirtschaftlicher Nutzung im Allgemeinen. Es lässt sich nicht abschließend klären, ob diese Features rein durch räumliche Kreuzkorrelationen mit dem Feature der Anbauflächen oder vielmehr komplementär zu den Anbauflächen auf Gebiete

mit PSM-Einsatz verweisen. Gleichzeitig konnte aufgrund der begrenzten Auswahl an Stoffen und ihren Eigenschaften als mobile und sehr mobile Stoffe nur ein kleiner Ausschnitt des Stoffspektrums in Einzelstoffmodellen beleuchtet werden. Wirkstoffmodelle und Wirkstoffgruppenmodelle konnten im begrenzten Rahmen getestet werden. Inwiefern die in den Metabolitenmodellen als relevant eingestuften Variablen Hinweise auf die Vulnerabilität eines Standortes geben können, wird im Folgenden betrachtet.

- ▶ *Lässt sich die Vulnerabilität von Grundwasser gegenüber PSM-Einträgen quantifizieren, klassifizieren und räumlich vorhersagen?*
- ▶ *Können die Ergebnisse in die Entwicklung eines bundesweiten Indexverfahrens zur Vulnerabilität des Grundwassers gegenüber PSM überführt werden?*

Daran schließen sich die folgenden Fragen an: Ist die Neuentwicklung eines Indexverfahrens auf Basis der Ergebnisse des Maschinellen Lernens möglich oder können existierende Ansätze weiterentwickelt werden? Und ist dafür eine stoffspezifische bzw. kulturspezifische Betrachtung sinnvoll bzw. notwendig?

Im Allgemeinen sind die in den Grundwasserproben gefundenen Konzentrationen das Ergebnis mehrerer sich überschneidender Prozesse. Daten zur tatsächlichen Applikation von Wirkstoffen stellen das wichtigste Bindeglied zwischen Landnutzung und Stoffkonzentrationen im Grundwasser dar. In Ermangelung dieser Information ergeben sich grundsätzliche Hürden. Im Grunde lassen sich zwei Fälle definieren. Entweder es handelt sich um:

FALL A: KONZENTRATION > LOQ AN WIRKSTOFF ODER METABOLIT WIRD GEFUNDEN.

Hierbei ist anzunehmen, dass eine Anwendung von PSM im Umkreis stattgefunden haben muss. Der Zeitpunkt und die Häufigkeit der Anwendung, die Distanz zur Messstelle und die konkreten, ausgebrachten Wirkstoffmengen sind hieraus nicht ersichtlich und unbekannt. Hier sei zur Vollständigkeit erwähnt, dass es neben der flächenhaften, fachgerechten Anwendung auch zu falsch-positiven Signalen kommen kann, was den Eintragspfad über flächenhafte Versickerung auf landwirtschaftlichen Flächen anbelangt. Das heißt, dass Konzentrationen im Grundwasser z.B. aus Punktquellen oder nicht fachgerechten Anwendungen (z.B. aus Unfällen) stammen können.

Ist der Wirkstoff/der Metabolit im Aquifer nachweisbar, weil der Standort sehr vulnerabel ist? Oder ist der Standort nur „mäßig“ vulnerabel und eine häufige Wirkstoff-Applikation fand statt? Wäre der Stoff häufiger appliziert worden, wären die nachweisbaren Konzentrationen höher? Aus der Tatsache, dass eine Stoffkonzentration im Grundwasser nachgewiesen werden konnte, lässt sich der Anteil der intrinsischen bzw. anwendungsbezogenen Vulnerabilität nicht zweifelsfrei und nicht ohne zusätzliche Informationen trennen. Die Quantifizierung der Vulnerabilität (stark, mäßig, schwach) ist daher vom Einfluss der tatsächlichen Anwendung schwer zu entkoppeln, wenn zu letzterer keine langjährigen Daten vorliegen. Die Unsicherheit in der Bewertung der „Stärke“ der Vulnerabilität liegt in der Diskrepanz begründet, die sich aus der Annahme potenzieller PSM-Anwendungsmengen und tatsächlich erfolgten Anwendungen ergibt.

FALL B: KONZENTRATION < LOQ AN WIRKSTOFF ODER METABOLIT WIRD NICHT GEFUNDEN.

Als Ursachen für Fall B kommen mehrere Möglichkeiten in Frage. Eine Option ist, dass die Anwendung tatsächlich nicht stattfand. Wenn die Anwendung nicht stattgefunden hat, zeigt sich hier das Problem der falsch-negativen Signale was den Eintragspfad über landwirtschaftliche Flächen betrifft. Die zweite Option ist, dass sich die Messstelle in einem relativ geschützten Aquifer befindet und der Stoff bislang nicht oder nicht in nachweisbaren Konzentrationen dort angekommen ist. Theoretisch ist es möglich, dass entsprechende Stoffkonzentrationen durch zu

hohe Bestimmungsgrenzen oder große Messintervalle in den Messprogrammen nicht erfasst werden. Geringe Konzentrationen sind folglich nicht zwangsläufig gleichbedeutend mit geringer intrinsischer Vulnerabilität. Zwar zeigten die Modellergebnisse eindeutig, dass sich die Kulturflächen in der Umgebung der Messstellen durchaus als Proxy für die Anwendungsmengen von PSM eignen, für eine konkrete Differenzierung zwischen den genannten Optionen ist für Fall B zum jetzigen Kenntnisstand die Historie der Anwendungsdaten erforderlich.

Aufgrund dieser Limitation sind der Bewertung der Grundwasservulnerabilität aus den vorliegenden Messdaten Grenzen gesetzt, insbesondere im Hinblick auf die Entwicklung eines Vulnerabilitätsindizes aus Messdaten. Idealerweise sollte ein Index ein ausreichendes Maß an Generalisierbarkeit aufweisen und nicht ausschließlich für einen bestimmten Stoff anhand bestimmter Messstellen zu bestimmten Zeitpunkten gelten, wofür ein gewisses Verständnis und ein gewisser Grad der Quantifizierung kausaler Zusammenhänge notwendig ist. Grundsätzlich ist es schwierig mittels statistischer Modelle, die aus retrospektiven Daten lernen, kausale Zusammenhänge zu erfassen. Dies liegt darin begründet, dass sich retrospektiv u.a. keine Störfaktoren (engl.: *confounding factors*) ausschließen lassen können. Solche Faktoren beeinflussen den kausalen Zusammenhang zwischen zwei Größen. Oftmals liegen hierzu unzureichende Daten vor oder diese Faktoren sind unbekannt. Zur Ableitung kausaler Zusammenhänge eignen sich kontrollierte Versuche mit möglichst randomisierten Bedingungen. Dies würde im konkreten Fall heißen, dass die Anwendung von Wirkstoffen beispielsweise gleichförmig über verschiedene Bodeneigenschaften erfolgt, sodass Korrelationen zwischen Bodeneigenschaften und Anbaugebieten der zugehörigen Kulturen vermieden werden. Grundsätzlich lernen die Modelle anhand der Daten statistische Zusammenhänge, die u.U. den kausalen Zusammenhängen so nah kommen, dass das Modell gut generalisiert und gut extrapoliert. Voraussetzung dafür ist, dass die wichtigsten Faktoren bekannt sind und die Datenlage eine ausreichende räumliche und zeitliche Auflösung aufweist.

Die Zielvariable der ML-Modelle (Stoffkonzentrationen bzw. -klassen) enthält somit immer jeweils einen Anteil der intrinsischen bzw. anwendungsspezifischen (kulturspezifischen) Vulnerabilität. In Ermangelung von realen Anwendungsdaten für die jeweiligen Kulturen als steuernde Größe der kulturspezifischen Vulnerabilität haben die ML-Modelle mit hoher Wahrscheinlichkeit räumliche Zusammenhänge erlernt, die auf wiederholter potenzieller Wirkstoffapplikation an den Messstellen beruht, gepaart mit einem großen Flächenanteil dieser jeweiligen Kultur im Pufferbereich der Messstelle. Daraus resultierend kann indirekt, also qualitativ auf die Applikation von Wirkstoffen geschlossen werden. Es ist jedoch anhand der potenziellen Anwendungen nicht möglich kausale, deterministische Modelle zu erarbeiten, die sich zur Analyse von Szenarien eignen. Diese Hypothese stand zu Beginn des Projektes im Raum. Es war zu prüfen, ob die ML-Modelle ohne die Kenntnis der Anwendungsdaten die Zusammenhänge zufriedenstellend wiedergeben, um die fehlende Kenntnis der Anwendungsdaten ausgleichen zu können. Die Ergebnisse unterstreichen, dass zum jetzigen Zeitpunkt die Anwendungsdaten zum Trainieren der ML-Modelle erforderlich sind.

Überlagert wird diese Erkenntnis dadurch, dass die Modelle Zusammenhänge lernen, die aus der räumlichen Verteilung der Messstellen, den unterschiedlichen Messprogrammen der Bundesländer sowie laborspezifischen Bestimmungsgrenzen resultieren.

Je nach Messstellenauswahl rücken andere räumliche Zusammenhänge in den Fokus. Wenn wir mit der Modellentwicklung und der Beschreibung der Vulnerabilität sichere Grundlagen der Risikobewertung erreichen möchten, müssen die tatsächlichen Aufwandsmengen die auf einem *Feld X* zum *Zeitpunkt Y* aufgebracht werden, bekannt sein. Die Bereitstellung dieser Daten würde den wissenschaftlichen Erkenntnisgewinn und die regulatorische Anwendungspraxis deutlich voranbringen.

Nichtdestotrotz wurde der Versuch unternommen die intrinsische von der anwendungsspezifischen Vulnerabilität zu trennen mittels Modellvorhersagen mit künstlich vergrößerten Anbauflächen an allen Messstellen. Es ist davon auszugehen, dass die intrinsische Vulnerabilität regional unterschiedlich gut erlernt worden ist. Die Zuverlässigkeit der räumlichen Vorhersage war jedoch zum Teil gering, was darauf schließt, dass sich das Modell nicht zur Testung von Szenarien eignet.

Letztendlich wird geschlussfolgert, dass sich anhand der vorliegenden Datengrundlage kein robuster Vulnerabilitätsindex entwickeln lässt. Daher sollte im nächsten Schritt geprüft werden, ob existierende Vulnerabilitätskonzepte anhand von Monitoringdaten evaluiert werden könnten. Zum einen wurde der Abgleich zwischen Schutzfunktion der Grundwasserüberdeckung (nach Hölting et al. (1995) und LAWA) und den Monitoringdaten und zum anderen zwischen GeoPELMO-Simulationsergebnissen und den Monitoringdaten durchgeführt. Es wird versucht mittels dieses Abgleichs den Informationsgehalt zur intrinsischen Vulnerabilität in den Messdaten besser einzuordnen.

Der Abgleich zwischen der Schutzfunktion der Grundwasserüberdeckung und den Konzentrationsklassen von Metaboliten an den Messstellen im FARM-Datensatz zeigte keinen Zusammenhang. Alleinig bei den Wirkstoffen zeigte sich eine Tendenz, dass Standorte mit „ungünstiger“ Schutzfunktion anteilig mehr Messstellen mit Überschreitungen aufweisen. Gleichzeitig muss erwähnt werden, dass an Standorten aller drei Kategorien der Schutzfunktion ein gewisser Prozentsatz an Messstellen mit definierten Konzentrationsüberschreitungen vorkommt. In der Kategorie „günstig“ sind Standorte vertreten, bei denen geschätzt wird, dass es mehr als 10 und z.T. mehr als 25 Jahre dauern kann, bis ein Stoff das Grundwasser erreicht. Bei den betrachteten Metaboliten handelt es sich ausschließlich um mobile und sehr mobile Metaboliten, die zum Teil von Wirkstoffen mit Jahre- und jahrzehntelanger Anwendungshistorie stammen. Es wird vermutet, dass die Beschreibung der intrinsischen Vulnerabilität (nach Hölting et al. 1995 und der LAWA) für diese Metaboliten mit hoher Mobilität, die von Altstoffen stammen, nicht hinreichend geeignet ist. Die Abbauprodukte haben mittlerweile zahlreiche Grundwasserleiter unterschiedlicher Kategorien erreicht und kommen ubiquitär vor. Gleichzeitig kann diskutiert werden, dass sich Messdaten dieser Stoffe aus diesem Grund vermutlich nicht eignen, um räumliche Unterschiede der intrinsischen Vulnerabilität zu erfassen. Die Abgleiche wurden versuchsweise auch für Stoffe, deren Zulassung weniger lange zurückliegend begann, sowie für Zeitspannen in der Vergangenheit durchgeführt. Es wurde damit keine bessere Übereinstimmung zwischen den Konzentrationsklassen und den Kategorien der Schutzfunktion der GW-Überdeckung festgestellt. Erschwerend kommt hinzu, dass bei neueren Stoffen (und vor allem ihren Metaboliten) oft eine gewisse Zeit verstreichen kann, bis sie in die Monitoringprogramme aufgenommen werden und Messwerte an einer ausreichenden Anzahl an Messstellen vorliegen. Somit konnten Vergleiche mit „jüngeren“ Stoffen nur im sehr begrenzten Maße getestet werden.

Die ML-Modelle für mobile und sehr mobile Metaboliten bzw. Wirkstoffe neigen dazu, die landwirtschaftliche Nutzfläche (konkrete Kultur oder Ackerland allgemein) als eines der wichtigsten, wenn nicht als das wichtigste Feature zu identifizieren. Hydrogeologische Eigenschaften spielen eine deutlich untergeordnete Rolle. Diese Beobachtung deckt sich mit der Abweichung zu den Schutzfunktionskategorien.

Bei den Wirkstoff- bzw. Wirkstoffgruppenmodellen war dies nicht immer der Fall. Gerade bei den moderat mobilen Wirkstoffen zeigten sich andere Feature (Boden, Hydrogeologie) auf den oberen Rangplätzen. Der Abgleich zwischen Hölting und den Überschreitungen der Wirkstoffe zeigte in der Tendenz, dass die Schutzfunktion der Grundwasserüberdeckung Wirkstofffunde besser beschreiben kann, bzw. dass die Wirkstoffe sich zur räumlichen Differenzierung von

intrinsischer Vulnerabilität ggf. besser eignen könnten als die untersuchten Metaboliten mit ihren Eigenschaften. Welche konkrete Rolle die Mobilität der Stoffe spielt, konnte nicht abschließend geklärt werden. Dies wird durch die ähnliche Feature-Gewichtung zwischen Metaboliten und mobilen Wirkstoffen unterstrichen. Auch die mobilen und sehr mobilen Wirkstoffmodelle zeigten eine geringe Feature-Gewichtung der hydrogeologischen Features während das Feature Ackerland als relativ wichtig eingestuft wurde. Selbstverständlich ist beim Vergleich zwischen Metaboliten und Wirkstoffen Vorsicht geboten, da letztere erst im Boden entstehen und dort ggf. länger gespeichert werden können. Wir vermuten, dass ggf. aus weniger mobilen Stoffen mehr über die intrinsische Vulnerabilität gelernt werden könnte. Gleichzeitig liegt es in der Natur der Sache, dass solche Stoffe und gerade Wirkstoffe in geringeren Konzentrationen auftauchen bzw. einen derart hohen Anteil zensierter Messungen aufweisen, dass daraus keine aussagekräftigen Abgleiche geschweige denn robuste Modelle trainiert werden können. Eine Verringerung und Harmonisierung der Bestimmungsgrenzen könnte dazu beitragen, bundesweite Wirkstofffunde zukünftig besser auswerten zu können.

Die Abgleiche zwischen Monitoringdaten und den prozessorientierten Modellergebnissen mit GeoPELMO DE zeigten größtenteils keine Übereinstimmungen. Grundsätzlich lässt sich beobachten, dass sich bei allen betrachteten Wirkstoffen und Metaboliten Messstellen mit Überschreitungen in allen räumlichen Perzentilklassen finden. Für Metazachlor ESA findet sich bei GeoPELMO DE (FOCUS L) der erwartete Zusammenhang (anteilig mehr Überschätzungen in der Perzentilkategorie 75-100), während bei GeoPELMO DE (FOCUS L+R+M+D) dies nicht klar ersichtlich ist. Für Metazachlorsäure zeigt sich bei beiden Modellkonfigurationen eine Tendenz zu anteilig mehr Überschätzungen in der Perzentilkategorie 75-100. Für Metolachlor ESA und Metolachlor OXA findet sich jedoch kein sichtbarer Zusammenhang.

Unter den Wirkstoffen findet sich bei Bentazon ein gegenläufiger Zusammenhang mit einem höheren Anteil der Messstellen mit Überschreitungen in der Perzentilkategorie 0-25 in beiden Modellkonfigurationen. Im Falle von Metolachlor, Metazachlor und Terbutylazin findet sich kein klarer Zusammenhang. Ein geringfügiger Zusammenhang kann höchstens bei Terbutylazin für die Modellkonfiguration GeoPELMO DE (FOCUS L+R+M+D) diskutiert werden.

Insgesamt lässt sich schlussfolgern, dass dieser Abgleich für die meisten Stoffe keine Übereinstimmung zwischen dem Anteil der Messstellen mit Überschreitungen und der höheren räumlichen Perzentilkategorie gezeigt hat. Es werden Messstellen mit Überschreitungen in allen räumlichen Perzentilklassen gefunden. Einen Vorteil von GeoPELMO DE (FOCUS L+R+M+D) mit zusätzlichen Prozessen gegenüber GeoPELMO DE (FOCUS L) lässt sich hieraus nicht ableiten. Das Vorhandensein von Makroporenfluss bei GeoPELMO zeigte bei Metaboliten nur Unterschiede von wenigen Prozent, wohingegen bei den Wirkstoffsimulationen die berechneten Konzentrationen zwei- bis zehnmal höher sein konnten (Klein et al. 2019; Klein et al. 2025). Der vorliegende Abgleich basiert somit größtenteils auf einer begrenzten Anzahl an Metaboliten, was die geringen Unterschiede zwischen den zwei Modellvarianten (GeoPELMO DE (FOCUS L) und GeoPELMO DE (FOCUS L+R+M+D)) erklären könnte. Aus diesem Grund ist eine vollumfängliche Einschätzung zum gegenwärtigen Zeitpunkt nicht möglich. Zusätzlich zu den Modellspezifika sehen wir weiterhin Verbesserungsmöglichkeiten bei der räumlichen Bodendatengrundlage, in welche die Umsetzung der Prozesse in den Simulationsroutinen eingebettet ist. Da der organische Kohlenstoffgehalt im Boden ein wichtiger Treiber der GeoPELMO-Modellroutinen ist, könnte hier die Auflösung der Inputdaten (Bodenkarten) eine entscheidende Rolle spielen. Möglich ist zudem, dass die Kombination aus kreisrundem 1 km-Radius um die Messstelle als auch die Auflösung der Bodenkarten (und somit auch C_{org}) die Gehalte des organischen Bodenkohlenstoffs unter Anbauflächen nicht ausreichend präzise getroffen wurden. Auch ist hervorzuheben, dass ein universeller Faktor zur Umrechnung

zwischen Humusgehalt und C_{org} von 1,72 eine starke Verallgemeinerung der realen Bedingungen ist und nur zutrifft, wenn der Anteil an C-Atomen der organischen Substanz 58% beträgt. Die organische Substanz des Bodens unterscheidet sich jedoch stark in ihrer chemischen Zusammensetzung in Abhängigkeit der Standortfaktoren und Bewirtschaftung und daher ist eine gewisse Unschärfe durch einen universellen Faktor zu erwarten. Neben der organischen Substanz sind der Anteil an Tonmineralen und der pH-Wert relevante Einflussfaktoren auf Sorptionsprozesse von PSM-Wirkstoffen und Metaboliten im Boden. Weiterhin spielen Redoxbedingungen sowie die Dynamik der Bodenfeuchte und -temperatur eine Rolle, die aufgrund der Auflösung der Bodendaten nicht ausreichend beschrieben werden kann. Auch wenn bisher viele Studien gezeigt haben, dass Sorptionsprozesse vorwiegend zwischen der organischen Substanz und Pestiziden stattfindet, so sind besonders quellfähige Tonminerale wie Smekтите in der Lage zu gleichem Anteil Atrazin zu binden wie die organische Substanz (Sheng et al. 2001). Die Wechselwirkungen von PSM-Wirkstoffen und deren Metaboliten in Abhängigkeit des pH-Wertes zu bewerten, ist äußerst komplex und uneinheitlich. Sorptionsisothermen für Diuron zeigen keinen Einfluss des pH-Wertes, wohingegen die Retentionseigenschaften von Bromoxynil und Ametryne stärker durch den pH beeinflusst werden (Sheng et al. 2005). Grundsätzlich kann eine Versauerung des Bodens (Absenkung des pH) oder eine Aufkalkung (Anhebung des pH) eine antagonistische Wirkung auf bestimmte PSM-Wirkstoffe ausüben (Arias-Estévez et al. 2008). Kausale Zusammenhänge für Wirkstoffgruppen zu identifizieren ist daher vermutlich schwierig zu realisieren. Ähnlich schwer zu messen und kausale Zusammenhänge herzustellen ist die Verwendung der Redoxbedingungen eines Standortes. Das Redoxpotential in wässriger Lösung gemessen ist abhängig von den beteiligten Redoxspezies (bspw. durch die Konzentrationen der Spezies Fe^{3+} und Fe^{2+}). Speziell im Grundwasser spielen Oxidationsprozesse eine Rolle, sobald eine Wasserprobe gefördert wird und in Kontakt mit O_2 kommt. Das Redoxpotential ist daher nicht statisch und sollte nicht als quantitative Größe betrachtet werden. Sowohl im Boden als auch im Grundwasser spielen thermodynamische Restriktionen eine Rolle, wodurch die Nernst-Gleichung als Grundlage zur Messung des Redoxpotentials (E_H) keine Anwendung findet und der E_H eher einen qualitativen bis semi-quantitativen Charakter hat (Whitfield 1974). Der E_H als Trainingsvariable für das ML-Modell hat daher vermutlich nur eine geringe Aussagekraft. Dennoch konnte in einem kleinskaligen Laborversuch gezeigt werden, dass unter nitratreduzierenden Bedingungen und Beteiligung von gelöstem organischen Kohlenstoff (DOC) 2,4-Dichlorphenoxyessigsäure besonders stark abgebaut wird. Ohne das Vorhandensein von DOC zeigten unter anoxischen Bedingungen die PSM Metaboliten wie 2,6-Dichlorobenzamide (BAM), Mecoprop-p (MCP) und der Wirkstoff Bentazon keine signifikanten Unterschiede in Abhängigkeit der Redoxbedingungen (Luo et al. 2019). Neben der Komplexität der genannten Prozesse und der Schwierigkeit ihrer Abbildung ist auch die Wahrscheinlichkeit der Ausbildung von Makroporen und ihre Auswirkungen auf quantifizierbare Stoffeinträge ins Grundwasser eng an die Datengrundlage geknüpft. Eventuell bietet sich ein Vergleich und eine Sensitivitätsanalyse von GeoPELMO-Simulationen an, die anhand von 5000er Bodenkarten durchgeführt werden und eine begrenzte Anzahl an räumlichen Pixeln betrachten. Damit könnte generell untersucht werden, inwiefern eine höhere Auflösung den Abgleich zwischen Simulations- und Monitoringergebnissen beeinflusst, ohne dabei einen unverhältnismäßigen Rechenaufwand zu betreiben. Hochaufgelöste Bodenkarten stehen nur für einen Teil der Bundesländer zur Verfügung. Wir empfehlen zudem, NZM-Standorte nicht anhand eines zu groben Modellmaßstabs abzuleiten und für solche standortbezogenen Modellierungen mit GeoPELMO DE ggfs. Bodendaten aus der BK1:5000 zu verwenden. Modelle, die EU-weit mit Eingangsdaten im Maßstab 1:1.000.000 arbeiten, sind daher mit Bezug auf die Vorhersagegüte der SDLMs als kritisch zu bewerten. Es bedarf weiterer

Anstrengungen, die erarbeiteten Modelle bzgl. ihrer Belastbarkeit in Sachen Vulnerabilität zu validieren.

Neben der Beschreibung der Bodeneigenschaften beinhalten die zusätzlichen Prozesse auch Informationen zu drainierten Flächen und zum Oberflächenabfluss. Während Drainagen und Abfluss den oberflächlichen Transport und räumliche Umverteilung von Stoffen durchaus beeinflussen, sind der aktuellen Datenverfügbarkeit Grenzen gesetzt. Bei der Wahrscheinlichkeit, dass ein Gebiet drainiert wird, handelt es sich um Schätzungen. Dokumentationen und Karten zu Drainagen liegen in der Regel nicht flächenhaft vor. Oft lässt sich gerade bei älteren Drainagen, die ggf. Lecks oder Verstopfungen aufweisen können, nicht direkt auf die tatsächliche dränierende Wirkung dieser Strukturen schließen. Nichtsdestotrotz hat eine Studie die besondere Sensitivität von PSM-Umverteilung nach Trockenphasen und nach Regenereignissen für Oberflächengewässer in kleineren (< 30 km²) Einzugsgebieten gezeigt (Halbach et al. 2021). Als maßgebliche Einflussfaktoren für das Auftreten von bis zu 76 Wirkstoffen und 32 Metaboliten in den Oberflächengewässern konnte die Halbwertszeit im Wasser und die Absatzmengen identifiziert werden. Dies unterstreicht die Bedeutung und Notwendigkeit, tatsächliche Aufwandmengen – sofern dies zukünftig möglich sein wird – mit in Betracht zu ziehen. Die Sensitivität für erhöhte Mobilität nach einem Regenereignis war für die Metaboliten geringer ausgeprägt als gegenüber den PSM-Wirkstoffen (Halbach et al. 2021).

Aus der Literatur lässt sich folglich eine starke Relevanz der genannten Einflussfaktoren für Transport- und Sorptionsprozesse ableiten, was die Notwendigkeit ihrer Berücksichtigung in Prozessmodellen unterstreicht. Zum gegenwärtigen Zeitpunkt ist es allerdings nicht möglich, die Ursachen für die Abweichung zwischen den Simulationen und den Monitoringdaten zweifelsfrei zu begründen – unabhängig davon, ob es sich um die Simulationen mit GeoPELMO DE (FOCUS L) oder GeoPELMO DE (FOCUS L+R+M+D) handelt. Grundsätzlich stellt sich die Frage, inwiefern die unterschiedlichen räumlichen und zeitlichen Skalen der simulierten Prozesse, die in der Bodenzone ablaufen, mit solchen Prozessen vergleichbar sind, die die Konzentration in Grundwasserproben beeinflussen. Sind Überschreitungen von 0,01, 0,05 bzw. 0,1 µg L⁻¹ mit räumlichen Perzentilklassen vergleichbar? Die Prozessmodelle nehmen eine räumlich gleichförmige, tägliche Applikation von Wirkstoffen an. Da tatsächliche Anwendungsdaten nicht bekannt sind, können Abweichungen zu Monitoring-Daten auch darin begründet sein, dass sich die tatsächlichen Applikationszyklen in dieser Annahme nicht widerspiegeln. Dennoch wird es als kritisch angesehen, dass die beiden verwendeten SDLM-Modellversionen von PELMO nicht in der Lage waren, einen bedeutenden Anteil an Messstellen bzw. Standorten mit nachweisbar höheren Wirkstoff- oder Metaboliteinträgen vorherzusagen. Eine Erklärung, dass es sich hierbei um falsch-positive Einträge handeln könnte, die nicht auf Versickerung zurückgehen könnten, greift hier möglicherweise zu kurz, weil ein Großteil der untersuchten Stoffe Metaboliten sind, die erst nach einer gewissen Zeit im Boden gebildet werden. Es könnte perspektivisch sinnvoll sein, sich von bundesweiten GeoPELMO-Simulationen im 1-km Pufferbereich der Messstellen zu lösen und stattdessen die Simulationen standort- und stoffspezifisch mit räumlich höher aufgelösten Inputdaten anzupassen und erneut zu berechnen. Der Nachteil einer solchen Vorgehensweise ist, dass ein bundesweiter Vergleich von quantitativer Vulnerabilität dadurch nicht möglich ist, sondern sich das Rechenergebnis immer nur auf einzelne Standorte bezieht. Die grundsätzliche Kritik, dass zwischen den Simulationen und den Monitoringdaten starke Skalenunterschiede den Vergleich erschweren, bezieht sich auf alle Prozesse (Transport, Sorption, Abbau).

Zusammenfassend lässt sich sagen, dass die Überlegungen zur Einordnung der ML-Modellergebnisse hinsichtlich ihrer Aussagekraft in der Vulnerabilitätsbewertung sowie der Abgleich der Index- und der prozessbasierten Verfahren mit den Monitoringdaten zahlreiche

Fragen aufwerfen. Beim betrachteten Index-Verfahren steht die Vermutung im Raum, dass diese bei Stoffen mit langer Anwendungshistorie und hoher Mobilität unzureichend sind. Die Abweichung zwischen den Simulationen aus Versickerungsmodellen und den Monitoringdaten zeigt in erster Linie die Skalenfrage auf: Sind die simulierten Sickerwasserkonzentrationen im Boden und die Konzentrationen im Grundwasser zeitlich und räumlich vergleichbar? Aus diesen Gründen lassen sich aus diesem Projekt keine klaren Empfehlungen ableiten, welches der beiden zur Verfügung stehenden Verfahren besser geeignet ist für eine verlässliche Vulnerabilitätsvorhersage.

6 Nachzulassungsmonitoring

6.1 Zielgerichtete Monitoringstudien im Nachzulassungsverfahren (NZM)

Einige der eingangs in Kapitel 1 dargelegten Ziele des Projektes beziehen sich auf die NZM-Studien. NZM-Studien als zielgerichtete Monitoringstudien unter bekannten Bedingungen dienen der Risikoüberprüfung, inwieweit konkrete PSM-Anwendungen zu Einträgen in den obersten Grundwasserleiter führen. Das erklärt hohe Anforderungen an die Messstellenauswahl, wie z.B. kurze Verweilzeiten von maximal wenigen Jahren. Derzeit werden solche Studien an wenigen (ca. 20) Messstellen durchgeführt. Die Messstellen bzw. die Standorte müssen daher eine hohe Vulnerabilität aufweisen, um sicherzustellen, dass sie der Überprüfung der bundesweiten Eintragsrisiken entsprechender PSM in Grundwasseraquifere standhalten. Aktueller Stand ist, dass sich für diese Studien vorwiegend Lockergesteinsgebiete mit freien Porengrundwasserleitern und sandigen Böden aufgrund ihrer hohen Durchlässigkeit eignen. Diese Standorte werden als besonders vulnerabel angesehen, da der Übergang von Wirkstoffen und Metaboliten in das Grundwasser innerhalb kurzer Zeiträume zu erwarten ist (König et al. 2020). Das stellt Anforderungen an die Grundwassermessstelle (GWMS), wie beispielsweise oberflächennahes Grundwasser und eine kurze Filterstrecke, die ebenfalls nah am Grundwasserspiegel liegt, ohne im Jahresgang trocken zu fallen. *Edge-of-field* Studien ermöglichen den Brückenschlag, wo die GWMS am Feldrand liegt und nach Applikation eines PSM-Wirkstoffes beprobt wird, um die Auswirkungen der PSM-Applikation eines Wirkstoffes auf die Vulnerabilität eines einzelnen Standortes zu untersuchen. Grundvoraussetzung dafür ist die Kenntnis über die Grundwasserfließrichtung und man muss mit dem Nachteil auskommen, dass Aussagen über größere Einzugsgebiete nicht möglich sind (Gimsing et al. 2019). Weiterhin wird die Frage gestellt, welche weiteren Faktoren im Sinne einer hohen Vulnerabilität bei der Standortauswahl zu bedenken sind. Ziel des FARM-Projektes war es, die Grundwasservulnerabilität auch hinsichtlich der weniger durchlässigen Substrate (Böden mit höheren Gehalten der Bodenartenklassen Schluff und Ton), die Makroporenbildung begünstigen, besser zu verstehen und ihre Eignung als NZM-Standorte abzuschätzen. Neben Porengrundwasserleitern sollte bewertet werden, ob sich auch Kluft- oder Karstaquifere, die vulnerabel gegenüber PSM-Einträgen sein können, als NZM-Messstellen eignen. Die an das NZM-Studiendesign gestellten Fragen sind damit Fragen zur intrinsischen Vulnerabilität.

Die Beantwortung dieser Fragen ist eng an die Bedingungen geknüpft, die zur Entwicklung eines Vulnerabilitätsindizes gegeben sein müssen. Wie in Kapitel 5.3 diskutiert, sind die fehlenden Informationen zu tatsächlichen PSM-Anwendungen ein entscheidender Faktor, um die intrinsische und die anwendungsbezogene Vulnerabilität eines Standortes differenzieren zu können. Da dies aufgrund der zur Verfügung stehenden Monitoringdatengrundlage nicht umsetzbar war, sind den Empfehlungen für die NZM-Standortauswahl Grenzen gesetzt.

Aus den Abgleichen zwischen Monitoringdaten und berechneten Erwartungskonzentrationen mittels GeoPELMO DE (FOCUS L bzw. FOCUS L+R+M+D) lassen sich hinsichtlich der Auswahl von Standorten für zielgerichtete NZM-Studien keine Empfehlungen für die Verwendung als NZM-Standorte ableiten. Denn die Vorhersagegüte der SDLM-Modelle zeigt, dass eine nicht unerhebliche Anzahl an Standorten als gering vulnerabel betrachtet wird, aber an den GWMS dennoch Wirkstoffe und Metaboliten in entsprechenden Konzentrationen gefunden wurden. Bei den betrachteten Metaboliten und Wirkstoffen konnte der Abgleich mit den Erwartungskonzentrationen die Monitoringdaten nicht widerspiegeln. Die Abweichungen lassen sich nicht eindeutig einem der möglichen Gründe, die zuvor diskutiert wurden, zuordnen. Bei den Abgleichen zwischen Monitoringdaten und der Schutzfunktion der

Grundwasserüberdeckung (nach Hölting et al. 1995 und LAWA) ließ sich für die betrachteten mobilen Metaboliten ebenfalls keine Übereinstimmung feststellen. Bei den Wirkstoffen und Wirkstoffgruppen zeigte sich allerdings, dass die anhand der Grundwasserüberdeckung als vulnerabel eingestuften Standorte einen geringfügig höheren Anteil an Messstellen mit Überschreitungen aufweisen. Gleichzeitig findet sich ein bestimmter Anteil an Messstellen mit Überschreitungen an Standorten, die als „geschützt“ eingestuft wurden. Anhand der Stoffe, für die eine ausreichende Datengrundlage für diese Abgleiche vorliegt, zeigen sich sowohl die beiden Versionen der GeoPELMO DE Modelle als auch die Schutzfunktion der Grundwasserüberdeckung als unzureichend, um ihr Vorkommen an Messstellen zu erklären. Es muss damit gerechnet werden, dass europäische SDLM-Modelle, die mit EU-weiten Eingangsdaten arbeiten noch ungenauer in ihrer Vorhersagegüte sind. Aus diesen Gründen ist zu empfehlen, bei der Auswahl von Standorten für zielgerichtete Monitoringstudien (NZM-Studien) derzeit vermehrt auf Expert-Judgement und nach Möglichkeit lokale Datenerhebungen zu setzen. Wir empfehlen weitere Anstrengungen zu unternehmen, die erarbeiteten Modelle bzgl. ihrer Treffsicherheit in Sachen Vulnerabilität zu validieren.

6.2 Messstellenauswahl nach Kriterien und Datenverfügbarkeit

Der Vorteil eines deutschlandweit harmonisierten Datensatzes der GWMS, an denen PSM-Wirkstoffe und deren Metaboliten gemessen werden, ist, dass eine Liste potenziell geeigneter Messstellen für zielgerichtete Monitoringstudien erstellt werden kann. Konkret können Standorte vorausgewählt werden, die einen Teil der geforderten Kriterien erfüllen (König et al. 2020). Die Kriterien zur Vorauswahl beziehen sich auf verfügbare Stammdaten der Messstelle, Grundwasser- und Bodengegebenheiten im Anstromgebiet:

- ▶ Grundwasserflurabstand,
- ▶ Filterlänge und -position,
- ▶ Grundwasserfließrichtung,
- ▶ Fließ- und Transportgeschwindigkeiten,
- ▶ Fundsituation der Wirkstoffe und Metaboliten,
- ▶ Art des Grundwasserleiters,
- ▶ Quellen,
- ▶ Anbau der Zielkultur,
- ▶ Punktquellen von PSM und
- ▶ Böden im Anstrom sowie Drainagen.

Wichtigstes Kriterium sind die aus einer Zeitreihe erhobenen Daten zum Grundwasserflurabstand einer Messstelle. Der mittlere Grundwasserflurabstand (GW_{mean}) sollte > 1 m und < 5 m unter Geländeoberkante (GOK) liegen. Der maximale Wert des Grundwasserflurabstandes (GW_{max}) sollte jedoch in keinem Fall tiefer als 10 m unter GOK liegen. Gebiete mit $GW_{\text{max}} > 10$ m unter GOK sind nur dann NZM-kompatibel, wenn nachgewiesen werden kann, dass die Verweilzeit in der vadosen Zone weniger als drei Jahre beträgt. Für die folgende Auswertung werden Angaben zur Verweilzeit nicht berücksichtigt. Weitere wichtige Kriterien sind die Position der Filteroberkante (FOK) und die Filterlänge. Die FOK sollte möglichst nah am GW_{mean} liegen. Toleriert werden Abstände bis zu 2 m (maximal 3 m) unterhalb

der Grundwasseroberfläche (GWO), ohne im Jahresgang trocken zu fallen. Eine Filterlänge bis zu 2 m (maximal 4 m) ist vorgegeben (Tabelle 6). Die Filterlänge errechnet sich dabei aus der Position der Filteroberkante abzüglich der Filterunterkante (FUK; Filterlänge = FOK – FUK). Daraus ergeben sich zwei Kategorien, die differenziert werden, eine strikte Auswahl mit höheren Anforderungen an die Messstelle und eine tolerante Auswahl (Tabelle 6).

Tabelle 6: Kriterien für die Auswahl von potentiellen Messstellen für das Nachzulassungsmonitoring (NZM).

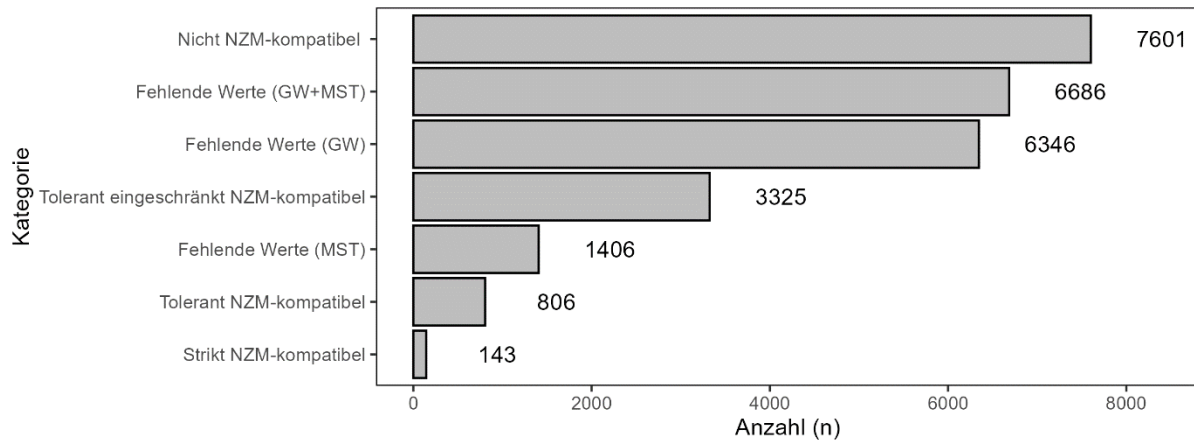
Kategorie	Grundwasserflurabstand	Filterposition	Filterlänge
Strikt	> 1 m bis 5 m im Mittel; nicht tiefer als 10 m	FOK auf Höhe GWO (ohne im Jahresgang trocken zu fallen), oder bis 2 m unterhalb GWS	Filterstrecke bis 2 m
Tolerant	> 1 m bis 5 m im Mittel; nicht tiefer als 10 m	FOK auf Höhe GWO (ohne im Jahresgang trocken zu fallen), oder bis 3 m unterhalb GWS	Filterstrecke bis 4 m

Es gibt bundeslandspezifische Unterschiede bei der Verfügbarkeit von Grundwasserstandsdaten (GW_{mean} , GW_{max}) und Stammdaten der Messstelle (FOK, FUK). Beide Informationen werden jedoch benötigt, um eine Vorauswahl geeigneter NZM-Messstellen zu treffen. Das Fehlen einer dieser Informationen hat automatisch zur Folge, dass eine Messstelle als nicht geeignet klassifiziert wird, obwohl die Messstelle eventuell NZM-kompatibel wäre. Es ist ersichtlich, dass es Unterschiede hinsichtlich der Datenverfügbarkeit gibt, wobei Informationen zum Grundwasserflurabstand häufiger ein fehlendes Kriterium sind als Informationen zur Filterposition und Filterstrecke. Es besteht jedoch generell die Möglichkeit, Informationen zum Grundwasserflurabstand nachzutragen, um eine Messstelle ggf. in die Vorauswahl mit einzubeziehen. In manchen Fällen fehlt nur die Information zur Filterunterkante, wodurch sich die Filterlänge nicht errechnen lässt. Diese sollte in einem NZM-Design möglichst kurzgehalten sein, bspw. nur 2 m Länge für das strikte Szenario und bis zu 4 m Länge für das tolerante Szenario. Die Bandbreite der verfügbaren Daten zur Vorauswahl NZM-kompatibler Messstellen variiert je nach Bundesland von nahezu vollständigen Informationen der benötigten Wertepaare bis hin zu dem vollständigen Fehlen zumindest einer benötigten Information. Geeignete Messstellen werden in Gebieten, wo alle Informationen vorhanden sind, häufig als kompatibel dargestellt. Abschließend lässt sich feststellen, dass neben den Stammdaten der individuellen Messstelle besonders die zeitliche Auflösung der Grundwasserstandsdaten eine Rolle spielt. Ohne das Kriterium des GW_{max} lässt sich nicht abschließend sagen, ob eine Messstelle im jahreszeitlichen Verlauf trockenfällt, also der Grundwasserflurabstand unter die Position der Filteroberkante absinkt. Dies ist ein Ausschlusskriterium für ein NZM-Design.

Von den ca. 26.000 Messstellen, die im FARM-Projekt untersucht wurden, sind 7.675 von den definierten Kriterien nicht NZM-kompatibel (Abbildung 121). Bei 6.686 Messstellen fehlten die Informationen zu Grundwasserstandsdaten und Informationen bezüglich Filterlänge und Position. Bei 6.377 Messstellen fehlten die Grundwasserstandsdaten und bei 1.407 Messstellen fehlte die Information zur Filteroberkante oder -unterkante. Die Kategorie „tolerant eingeschränkt NZM-kompatibel“ wurde gesondert ausgewiesen, da hier die Informationen zum Grundwasserflurabstand fehlen, aber sonstige Informationen zur Messstelle vorhanden waren unter dem Kriterium der Filterstrecke von 4 m Länge. Dies ist eine realistische Annahme, da der Ort der Beurteilung häufig die Grundwasseroberfläche ist. Es ist daher anzunehmen, dass eine Filteroberkante, die oberflächennah anzutreffen ist, ebenfalls oberflächennahes Grundwasser aufweist. Bei Bedarf und Hinzunahme von nachgereichten Informationen zum Grundwasserflurabstand könnten die Messstellen mit der Kategorie „tolerant eingeschränkt

NZM-kompatibel“ also potenziell als Untersuchungsstandorte berücksichtigt werden. Abschließend gibt es 806 Messstellen, die derzeit alle Anforderungen für die tolerante Kategorie erfüllen, wovon 143 Messstellen auch die strikten Anforderungen erfüllen.

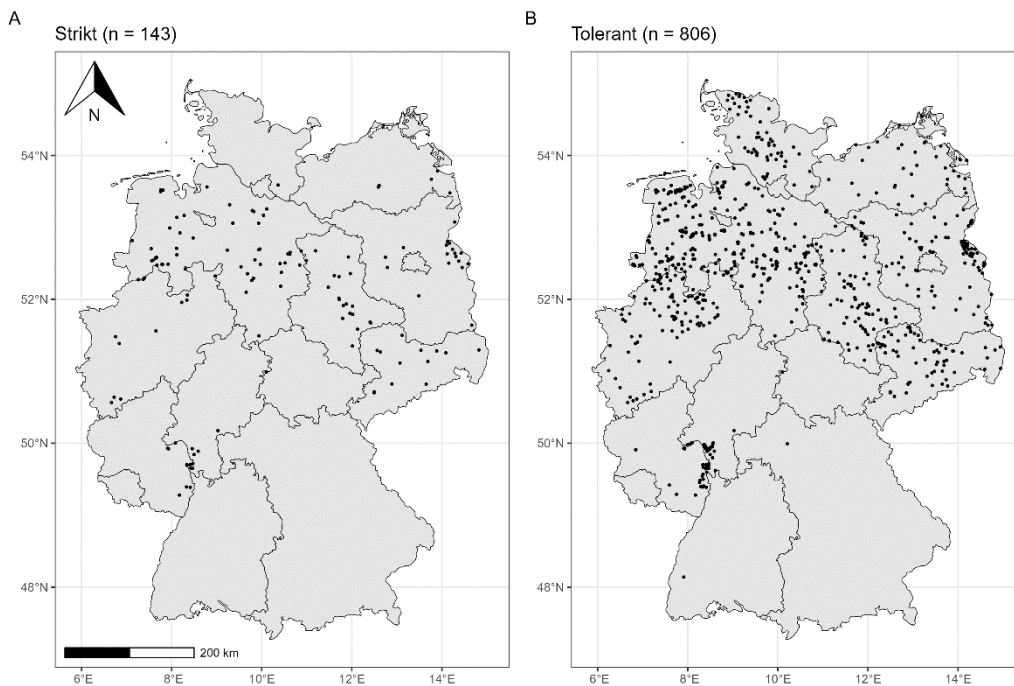
Abbildung 121: Kategorien von Messstellen aus dem FARM-Projekt die für das Nachzulassungsmonitoring (NZM) relevant oder nicht relevant sind.



Quelle: eigene Darstellung BGR.

Aus dem Datensatz wurde durch die Vorauswahl die Anzahl der Messstellen von 26.000 auf 806 reduziert. Dies entspricht einem Anteil von 3%. Einige Kriterien, die nicht in der FARM-Datenbank erfasst sind, müssen ggf. später gesondert evaluiert werden (bspw. Grundwasserfließrichtung, Fließ- und Transportgeschwindigkeiten, Böden im Anstrom etc.). Die reduzierte Auswahl muss vor einer Verwendung in einer NZM-Studie zudem auf weitere Kriterien hin untersucht werden (bspw. Anbau der jeweiligen Zielkultur, je nach zu untersuchendem Wirkstoff oder Metabolit). Die mit Abstand meisten potenziell geeigneten Messstellen liegen in Niedersachsen, gefolgt von Brandenburg und Nordrhein-Westfalen. Die NZM-kompatiblen Messstellen liegen vorwiegend im Norddeutschen Tiefland (Abbildung 122). In den Bundesländern Rheinland-Pfalz, Hessen, Thüringen, Saarland, Baden-Württemberg und Bayern kommen nur vereinzelt Messstellen infrage. Den limitierenden Faktor bildet hier vor allem die Filterstrecke, die im Norddeutschen Tiefland sowie im nördlichen Teil des Oberrheinischen Tieflands die entsprechende Länge aufweist, während in den topographisch höher gelegenen Regionen zumeist längere Filter verbaut wurden. Besonders im Norddeutschen Tiefland erfüllt der Großteil der Messstellen mit einer Filterlänge bis zu 2 m die strikten Kriterien für das NZM. Im Rhein-Main-Gebiet gibt es eine Häufung an Messstellen, die die toleranten Kriterien für die Filterlänge bis 4 m erfüllen. Besonders in Bayern und dem Saarland schließen tiefere Grundwasseroberflächen eine NZM-Nutzung aus (Abbildung 123).

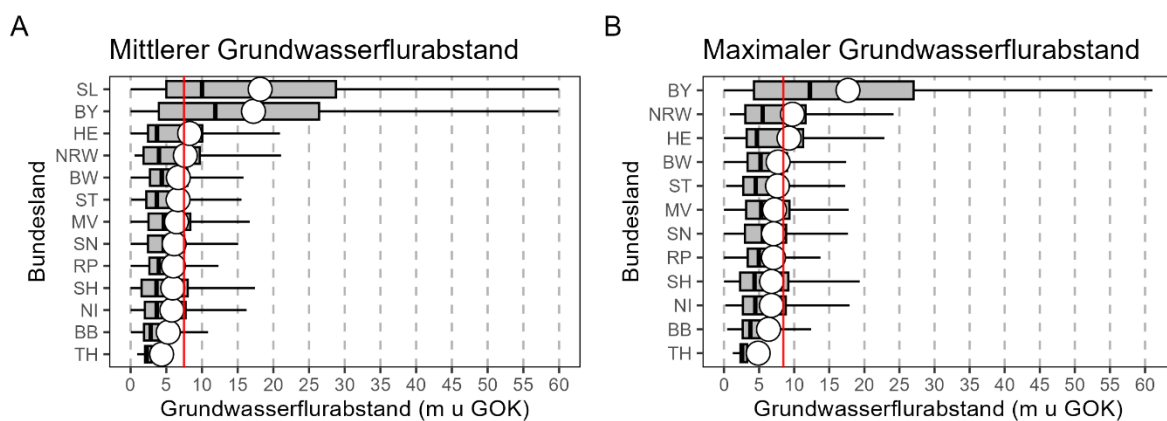
Abbildung 122: Lage der NZM-kompatiblen Messstellen des FARM-Projektes differenziert nach strikten (A) und toleranten (B) Kriterien.



Quelle: eigene Darstellung BGR.

Die Kriterien in strikt und tolerant ergeben sich aus Grundwasserstandsdaten und Messstelleninformationen (Filterposition, Filterstrecke). Die Kartenansicht zeigt deutlich, dass der Großteil der potenziell geeigneten Messstellen im Norddeutschen Tiefland liegt. Einige Messstellen liegen im Rhein-Main-Gebiet.

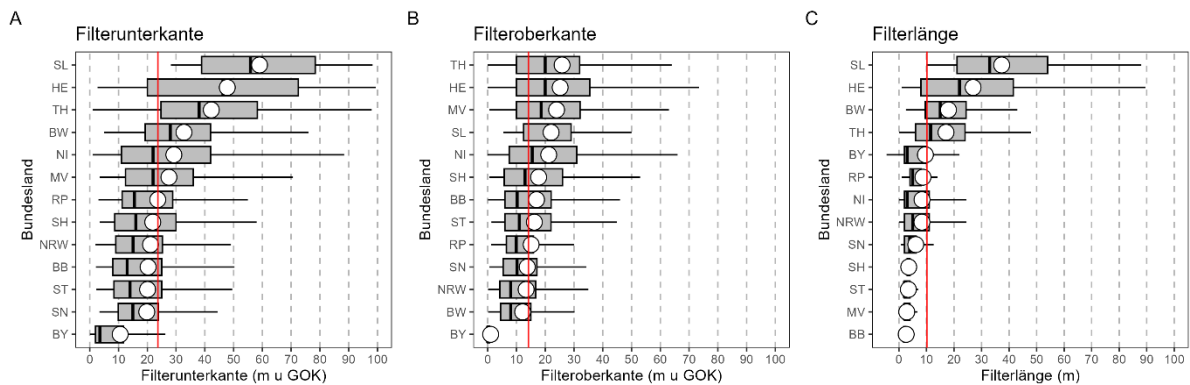
Abbildung 123: Statistische Kennzahlen zum mittleren (A) und maximalen Grundwasserflurabstand (B) je Bundesland.



Quelle: eigene Darstellung BGR.

Die graue Box kennzeichnet das 25te und 75te Perzentil der Daten, der schwarze Strich in der grauen Box ist der Median und der weiße Punkt ist der Mittelwert. Die rote Linie ist der Mittelwert aller Messstellen. Die Darstellung bezieht sich ausschließlich auf die NZM-kompatiblen Messstellen (n = 806).

Abbildung 124: Statistische Kennzahlen der Metriken zur Filterunterkante (A), Filteroberkante (B) und Filterlänge (C) je Bundesland.

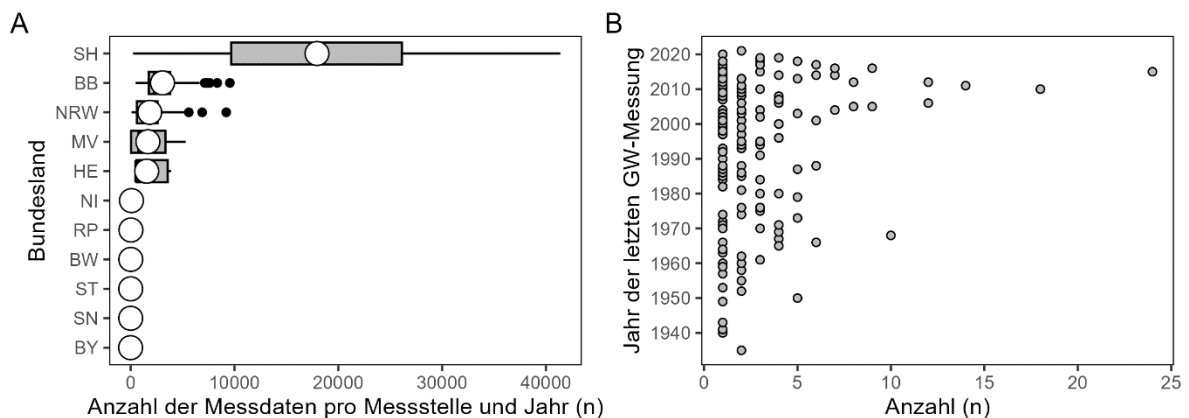


Quelle: eigene Darstellung BGR.

Die graue Box kennzeichnet das 25te und 75te Perzentil der Daten, der schwarze Strich in der grauen Box ist der Median und der weiße Punkt ist der Mittelwert. Die rote Linie ist der Mittelwert aller Messstellen. Die Darstellung bezieht sich ausschließlich auf die NZM-kompatiblen Messstellen (n = 806).

Die Auswahl der Messstellen hat aufgrund der heterogenen Datenlage der herangezogenen Kriterien eine variierende Zuverlässigkeit. Besonders die Messung des Grundwasserflurabstands erfolgt in den Bundesländern in unterschiedlichen Intervallen. Insbesondere wenn lediglich eine Grundwasserstandsmessung pro Jahr vorliegt, ist ein Trockenfallen des Filters im Jahresgang nicht auszuschließen. In Niedersachsen ist die Datenlage etwas besser mit rund zehn Messungen pro Jahr. In Thüringen und Bayern liegen rund 50 Messungen pro Jahr vor; in Mecklenburg-Vorpommern, Brandenburg, Nordrhein-Westfalen und Hessen 50 bis teilweise einige hundert. Die höchste Datendichte besteht in Schleswig-Holstein mit bis zu 20.000 Messungen pro Jahr (dies entspricht einem halbstündlichen Messintervall) (Abbildung 125A). Bei der Auswahl von Messstellen für das NZM wurden Messstellen ausgeschlossen, die nachweislich im Jahresgang trockengefallen sind. Es lässt sich (vor allem in den Bundesländern mit nur einer Messung) nicht ausschließen, dass davon noch weitere der 806 Messstellen betroffen sind. Ein weiterer Punkt, den man berücksichtigen muss, ist die Aktualität der letzten Messung, für die Grundwasserstandsdaten vorliegen und verwertet wurden. Die ältesten Informationen stammen aus den 1950er Jahren und sind daher von einer sehr geringen Aussagekraft (Abbildung 125B). Das aktuellste Datum der NZM-kompatiblen Messstellen stammt aus dem Jahr 2022 wobei bis zu 170 der 806 Messstellen als letzten Zeitstempel das Jahr 2010 aufweisen.

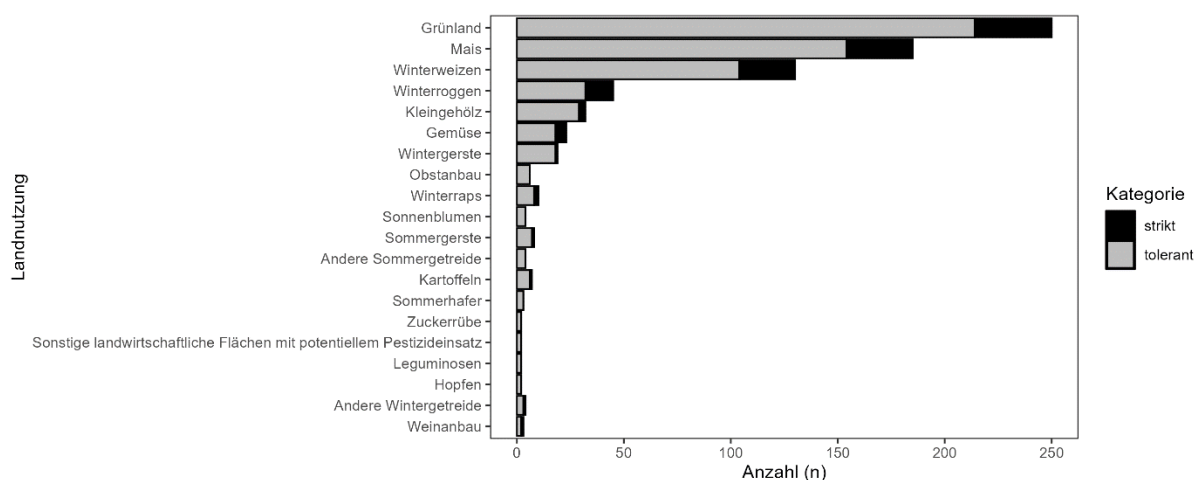
Abbildung 125: Informationen zur Messhäufigkeit des Grundwasserflurabstandes je Messstelle und Bundesland (A) und Jahr der letzten Messung, in der die Messung durchgeführt wurde (B).



Quelle: eigene Darstellung BGR.

Die flächendominierende Gesteinsart im Puffer ist bei 98% der (tolerant) ausgewählten Messstellen Sedimentgestein. Bei 91% der Messstellen liegt ein Porengrundwasserleiter vor (präferierte Hohlraumart für NZM). Rund 8% sind Kluftgrundwasserleiter. Die restlichen Messstellen fallen in die Klassen Kluft-/Poren- bzw. Kluft-/Karstgrundwasserleiter. Für eine engere Auswahl von Messstellen für das NZM sind die angebauten Kulturen entscheidend, damit Messstellen ausgewählt werden können, an denen der zu messende Stoff angewendet wird. An den vorausgewählten 806 Messstellen machen 21 der 27 Kulturklassen mindestens einmal den größten Flächenanteil im Puffer aus (Abbildung 126). Dabei stellen besonders oft die Klassen Grünland (n = 214), Mais (n = 154) und Winterweizen (n = 104) den größten Anteil. Zudem gibt es Klassen, die zwar nicht den größten Flächenanteil stellen, jedoch an vielen Messstellen einen erheblichen Anteil ausmachen. Dazu gehören Winterroggen, Wintergerste, Winterraps, andere Wintergetreide, Zuckerrübe, Gemüse und Kartoffeln. Besonders Obstanbau, Weinanbau und Kleingehölze kommen eher vereinzelt und kleinflächig vor (Abbildung 126).

Abbildung 126: Häufigkeit der dominierenden Landnutzung innerhalb des Puffers einer NZM-kompatiblen Messstelle abgeleitet aus Sentinel-Daten differenziert für die strikte und tolerante Kategorie.



Quelle: eigene Darstellung BGR.

Die Analyse zeigt, dass sich potentiell 143 (mit strikten Kriterien), bzw. 806 Messstellen (mit toleranten Kriterien) aus der FARM-Datenbank für NZM-Studien eignen könnten. Dies bezieht sich auf die Vorauswahl hinsichtlich der Informationen zum Grundwasserflurabstand und Informationen des Filterrohres. Es müssen weiterhin andere Kriterien erfüllt sein, bevor eine Messstelle die Kriterien des derzeit empfohlenen NZM-Designs erfüllt. Die drastische Reduktion der ca. 26.000 Messstellen, für die PSM-Daten der Wirkstoffe und Metabolite vorliegen, geht zum einen auf die reduzierte Datenverfügbarkeit von Grundwasserstandsmessungen (nur für ca. 6.400 Messstellen vorhanden) und zum anderen auf die Anforderungen an die Filterstrecke (die in topographisch höher gelegenen Regionen oft zu lang ist) zurück. Durch die Abwägung weiterer Parameter, die nicht in der Datenbank enthalten sind (z.B. Anstromrichtung, Punktquellen von Wirkstoffen etc.), sowie ggf. durch den Ausschluss von Messstellen mit einer unzureichenden Frequenz an Grundwasserstandsmessungen, kann sich die Anzahl geeigneter Messstellen noch weiter reduzieren. Die ausgewählten Messstellen liegen überwiegend im Norddeutschen Tiefland und teilweise im Rhein-Main-Gebiet. Die angebauten Kulturen im Puffer sind sehr variabel (alle Kulturklassen außer Hopfen sind vorhanden), sodass potentiell ein breites Spektrum an Wirkstoffen untersucht werden könnte. Abschließend sei erwähnt, dass der Zusammenhang der NZM-kompatiblen Messstellen mit der Landnutzung aus einem 1-km² großen Puffer um die jeweilige GWMS abgeleitet wurde. Diese Informationen können nur für eine erste Annäherung an die Frage der Landnutzung verwendet werden, weil in einer NZM-Studie letztlich nur die Felder im direkten Grundwasseranstrom betrachtet werden. Dieser Anforderung konnte in der vorliegenden Auswertung aufgrund deutschlandweit fehlender einheitlicher Informationen zur Grundwasserfließrichtung nicht Rechnung getragen werden.

6.3 Empfehlungen für zielgerichtete Monitoringstudien

Einige der eingangs in Kapitel 1 dargelegten Ziele des Projektes beziehen sich auf NZM-Studien, die innerhalb des Zulassungsverfahrens von PSM durchgeführt und bewertet werden. NZM-Studien dienen der Risikoüberprüfung, inwieweit konkrete PSM-Anwendungen zu Einträgen in den obersten Grundwasserleiter führen, was hohe Anforderungen an die Messstellenauswahl, wie z.B. kurze Verweilzeiten von maximal wenigen Jahren, erklärt.

Das heißt, dass grundsätzlich zwei Fragen beantwortet werden sollten:

- 1) Welche Standortbedingungen müssen vorliegen, damit Messstellen für das NZM in Frage kommen? Hierbei wurde explizit die Frage gestellt, ob die derzeit empfohlenen naturräumlichen Eigenschaften (bspw. sandige Bodenarten) ausreichend sind, oder ob Standorte mit anderen Bedingungen dafür geeignet sind.
- 2) Wie viele Messstellen werden aus statistischer Sicht im NZM für robuste Aussagen benötigt?

Aktueller Stand ist, dass für diese Studien vorwiegend Lockergesteinsgebiete mit freien Porengrundwasserleitern und sandigen Böden aufgrund ihrer hohen Durchlässigkeit als geeignet angesehen werden (König et al. 2020). Grundsätzlich wird die Frage gestellt, welche Faktoren bei der Standortauswahl zu bedenken sind. Ziel des Projektes war es, die Grundwasservulnerabilität hinsichtlich schluff- und tonreicher Böden bzw. Substrate, die Makroporenbildung begünstigen, besser zu verstehen und ihre Eignung als NZM-Standorte abzuschätzen. Neben Porengrundwasserleitern sollte bewertet werden, wie vulnerabel Kluff- bzw. Karstaquifere gegenüber PSM-Einträgen sind, und ob solche Standorte für NZM-Studien geeignet sind. Die an das NZM-Studiendesign gestellten Fragen sind damit Fragen zur intrinsischen Vulnerabilität.

Dieser Frage kann sich mittels multivariater Statistik (z.B. Clusteranalyse, Hauptkomponentenanalyse), Geostatistik (z.B. Kriging) oder maschinellem Lernen genähert werden. Mit dem Ziel, die Einflussfaktoren auf die Grundwasserqualität anhand der Daten zu

identifizieren, wurden in diesem Projekt Random Forest und Neuronale Netze eingesetzt. Voraussetzung zur Auswahl als NZM-Messstelle ist eine hohe Vulnerabilität. Die in diesem Bericht beschriebenen Modelle identifizierten die Landnutzung und die Filtertiefe (definiert als Filteroberkante) als relevante Variablen, konnten aber anhand der verwendeten hydrogeologischen Features hinsichtlich der intrinsischen Vulnerabilität für mobile Metaboliten keine verlässlichen Aussagen treffen.

Damit ist die Beantwortung dieser Fragen eng an die Bedingungen geknüpft, die zur Entwicklung eines Vulnerabilitätsindex gegeben sein müssen. Wie in Kapitel 5.3 diskutiert, sind die fehlenden Informationen zu tatsächlichen PSM-Anwendungen ein entscheidender Faktor, um die intrinsische und die anwendungsbezogene Vulnerabilität eines Standortes differenzieren zu können. Da dies in dem durchgeführten Projekt mit Restriktionen belegt ist, sind den Empfehlungen für die NZM-Standortwahl Grenzen gesetzt. Aus den ML-Modellen für die Metaboliten konnte kein starker Einfluss der Features Aquifer-kf, Hohlraumart und Makroporenklasse abgeleitet werden. Wie in Kapitel 4 diskutiert, ist davon auszugehen, dass die Modelle die Konzentrationsklasse vorwiegend über die Filtertiefe und die Flächengröße der relevanten Kulturen im Puffer abschätzen. Weitere Standortfeatures fließen größtenteils aufgrund ihrer Eigenschaft als Zeiger für landwirtschaftliche Flächen in das Modell ein. Gleichzeitig wurde kein Zusammenhang zwischen der Schutzfunktion der Grundwasserüberdeckung und den Konzentrationsklassen der Metaboliten gefunden. Die gesuchten Zusammenhänge zwischen den Stoffkonzentrationen und den Features der intrinsischen Vulnerabilität konnten aus den Metaboliten-Modellen nicht abgeleitet werden. Hierbei spielt eine Rolle, dass es sich um mobile Stoffe mit langer Anwendungshistorie handelt, wodurch diese Stoffe in der Zwischenzeit zahlreiche unterschiedliche Aquifere erreicht haben. Vereinzelt zeigt sich, dass Karst- und Kluftgrundwasserleiter bei manchen Stoffen leichte Unterschiede der Feature-Effekte aufweisen. Die Differenzierung der Prozesseffekte (Stofftransport, Versickerungsverhalten) und des Einflusses der Messstellenverteilung, des Monitoringdesigns und der Landnutzung ist ohne PSM-Anwendungsdaten jedoch kaum möglich. Ausgehend von diesen Modell-Ergebnissen konnte daher nicht abschließend geklärt werden, ob Studien in Lockergestein und sandigen Böden ausreichend sind, oder ob ergänzend Kluftgrundwasserleiter und bindige Böden ergänzend zu untersuchen sind. Gleichzeitig zeigt die Fundsituation von Wirkstoffen und Metaboliten, dass nicht alleinig sandige Standorte auffällig werden. Dies ist ein Hinweis darauf, dass bindige Böden berücksichtigt werden sollten, obwohl ihr Einfluss anhand der im Projekt verwendeten Methoden und Datengrundlagen nicht abschließend quantifiziert werden konnte. Ein adaptiertes Studiendesign sollte sich vermutlich an den Standorteigenschaften im Anstrom einer Messstelle richten. Sind die Grundwasserschwankungen aus einer retrospektiven Datengrundlage bekannt, empfiehlt es sich, nicht mit einer universellen Filtertiefe zu arbeiten. Um die jahreszeitlichen Schwankungen des Grundwasserflurabstandes zu berücksichtigen, sollten Messstellen mit unterschiedlichen Positionen der Filteroberkante verwendet werden. Ist eine ausgeprägte Heterogenität der Böden und der hydrogeologischen Bedingungen im Anstrom der Grundwassermessstelle zu erwarten, wird eine höhere Anzahl der Messstellen pro Standort empfohlen.

Funde von Wirkstoffen scheinen gegenüber Metaboliten eher dazu geeignet, die intrinsische Vulnerabilität eines Standortes zu beschreiben, wie die Abgleiche mit der Schutzfunktion der Grundwasserüberdeckung in Teilen vermuten lassen. Bei den ML-Modellen der moderat mobilen Wirkstoffe wird deutlich, dass eine hohe hydraulische Leitfähigkeit des Aquifers und das Vorhandensein von Karstaquiferen auf eine höhere Wahrscheinlichkeit der höheren Konzentrationsklasse verweist. Dies spricht dafür, dass sich mit Wirkstoffen und insbesondere Wirkstoffen, die nicht sehr mobil sind, ggf. mehr über die intrinsische Vulnerabilität lernen ließe. Die Problematik der fehlenden Anwendungsdaten bleibt hier allerdings bestehen. Gleichzeitig

sind der Stoffauswahl hier Grenzen gesetzt, weshalb dieses Potenzial nicht ausgeschöpft werden konnte. Die gezeigten Einzelstoffmodelle beschränken sich auf Altstoffe. Ein Großteil der Wirkstoffe eignet sich aufgrund des hohen Anteils linkszensierter Messdaten nicht für Einzelstoffmodelle. Darüber hinaus scheitern Einzelstoffmodelle für Wirkstoffe mit einer kürzeren und jüngeren Zulassungshistorie an der Verfügbarkeit von Monitoringdaten. Es ist davon auszugehen, dass das Vorhandensein von tatsächlichen Anwendungsdaten dazu beitragen kann, den Anteil der intrinsischen Vulnerabilität mittels ML-Modell verlässlicher zu differenzieren. Allerdings könnte ein Mehrwert erst entstehen, wenn Monitoringdaten von weiteren PSM-Stoffen und Stoffgruppen nutzbar wären.

Aus den ML-Modellergebnissen und den Abgleichen zwischen weiteren Verfahren zur Vulnerabilitätsabschätzung konnten wir dennoch Empfehlungen hinsichtlich der Datengrundlage für zielgerichtete Monitoringstudien ableiten. Wie die Modelle mit der erweiterten Feature-Auswahl gezeigt haben, könnte die GW-Chemie als räumlicher Zeiger für geeignete Standorte dienen. Vor allem Nitrat- und Kaliumkonzentrationen an der Messstelle können u.a. einen Hinweis auf die Intensität der landwirtschaftlichen Nutzung im Einzugsgebiet geben. Weiterhin wird die Verwendung von InVeKoS- oder Fernerkundungsdaten zur Identifikation von Anbaugebieten und der Vorauswahl von möglichen Messstellen bzw. Standorten empfohlen, was einen Vorteil gegenüber der Verwendung von aggregierten Agrarstatistiken auf Kreisebene mit sich bringt. Aufgrund von Unsicherheiten in den Modellen sind Bodenkarten mit höherer Auflösung (z.B. Bodenkarte BK 1:5000 anstelle der im Modell genutzten Bodenübersichtskarte (BÜK 1:250.000) und lokale Datenerhebungen für Studien vor Ort notwendig. Dabei handelt es sich um Beprobungen und Boden-Laboranalysen der Ackerflächen im direkten Anstrom der Messstelle.

Es ist sinnvoll sich der Beschreibung der intrinsischen Vulnerabilität von Kluft- und Karstaquiferen mittels anderer Verfahren und veröffentlichter Studien (z.B. gezielte Tracer-Versuche) zu nähern und ggf. in ausgewählten Gebieten Messungen von Wirkstoff- bzw. Metabolitenkonzentrationen in angepassten Messkampagnen durchzuführen. Sind solche Standorte aufgrund anderer Studien bereits als vulnerabel bekannt und sind die Verweilzeiten kurz genug für NZM-Studien, dann sollten diese mit existierenden Wirkstoff- und Metabolitenmessungen abgeglichen werden. In Gebieten, in denen lokale Informationen zu Fließwegen und Stofftransport vorliegen, kann ergänzend mit numerischen Grundwassermodellen wie MODFLOW/MT3DMS gearbeitet werden. Dies wird jedoch nur an gut untersuchten und instrumentell ausgestatteten Versuchsstandorten möglich sein, da eine Vielzahl hochqualitativer Inputdaten notwendig ist, um Parametrisierung und Randbedingungen auf einer kleinen Skala passend abzudecken. Darüber hinaus sind Informationen zur Anstromrichtung und Grundwassergleichenpläne bundesweit in sehr unterschiedlicher Qualität und Flächendeckung vorhanden, sodass wir für die Auswahl von NZM-Messstellen empfehlen, sowohl bei Karst- und Kluftaquiferen als auch bei Porengrundwasserleitern, Untersuchungen der lokalen Gegebenheiten vorzunehmen. Die Bestimmung der Fließrichtung des Grundwassers in Karstgebieten ist besonders anspruchsvoll. In der Regel muss auf eine Kombination aus verschiedenen Methoden zurückgegriffen werden, die u.a. Tracer-Versuche und Untersuchungen von Quellschüttungen (Eigenschaften des Wassers um dessen Herkunft zu unterscheiden), geophysikalische Charakterisierung von Karsthohlräumen sowie hydraulische Messungen umfassen, wobei letztere nur begrenzt aussagekräftig sein können. Für die Bewertung eines Standortes lassen sich Karst- und Kluftaquifere nur deutlich aufwendiger für ein NZM-Design anwenden, was nicht bedeutet, dass diese Standorte nicht weniger wichtig für die Vulnerabilitätsbewertung des Grundwassers sind. Grundwasserschwankungen unterliegen hier größeren Amplituden und die Wahl einer geeigneten Filtertiefe und Strecke gestaltet sich schwieriger gegenüber einem NZM-Design in einem Porengrundwasserleiter. Die

Randbedingungen zu hydraulischen Leitfähigkeiten und die Anstromrichtung lassen sich bei einem Porengrundwasserleiter besser fassen und quantifizieren, was die Verbindung zwischen dem Einsatz eines PSM-Wirkstoffes auf einer Fläche mit dem Signal im Grundwasser ermöglicht.

Die Ergebnisse aus diesem Projekt zeigen, dass sich den Fragen zur NZM-Standortauswahl nicht ausschließlich mittels der vorliegenden Monitoringdaten genähert werden konnte und es zu ihrer Beantwortung weiterer Daten und anderer Ansätze bedarf. Aufgrund der beschriebenen lokalen Komplexität von Karstaquiferen sind bundesweite ML-Modelle nicht in der Lage hierzu verlässliche Einschätzungen zu gewährleisten. Anhand der Funde, die in Karst- und Kluftaquiferen vorkommen (Cooke et al. 2024), lässt sich schlussfolgern, dass diese Gebiete auffällig werden können und es lohnenswert ist, nach Möglichkeiten zu suchen um sie hinsichtlich NZM berücksichtigen zu können. In sandigen Porengrundwasserleitern lassen sich NZM-Bedingungen jedoch weitaus besser prüfen. Aus den vorliegenden Projektergebnissen allein lässt sich nicht ableiten, inwiefern Gebiete mit Kluft-/Karstquiferen im Vergleich zu Porengrundwasserleitern gegenüber Wirkstoffen und Metaboliten von Pflanzenschutzmitteln vulnerabler sind oder nicht. Zum gegenwärtigen Zeitpunkt stehen nicht genügend aussagekräftige Informationen zur Verfügung, um Kluft- und Karstgebiete aufgrund der Schwierigkeit, den Stoffeintrag an einer Messstelle bestimmten einzelnen Flächen zuzuordnen zu können, sinnvoll in NZM-Studien zu integrieren. Es bedarf grundsätzlich mehr Daten, Analysen und langjährige Kenntnis zum Stofftransport um Kluft-/Karstquifere für einzelne NZM-Studien zu nutzen, vermutlich mit einem sehr standortspezifischen Studiendesign.

Um sich der zweiten Frage nach der notwendigen Anzahl an Messstellen zu nähern, gibt es eine Reihe an hilfreichen Ansätzen, die im Folgenden lediglich perspektivisch beschrieben werden können. Von ihrer Anwendung wurde abgesehen, da die Projektergebnisse keine belastbare Beschreibung der intrinsischen Vulnerabilität von Standorten liefern konnten.

Steht eine Auswahl an Messstellen bereit, welche die Bedingungen für das NZM-Monitoring erfüllen und denen eine gewisse intrinsische Vulnerabilität aufgrund der zuvor beschriebenen lokalen Untersuchungen zugeschrieben wird, kann die notwendige Mindestanzahl an Messstellen mittels Messnetzoptimierung berechnet werden. Die Entscheidung darüber, wie viele Messstellen erforderlich sind und welche Standorte in Frage kommen, kann mittels verschiedener statistischer Methoden erfolgen, verbunden mit zahlreichen Anwendungsbeispielen im Kontext von Grundwasserquantität und -qualität (Ohmer et al. 2019; Ohmer et al. 2022; Zhang et al. 2005). Mathematische Optimierungsverfahren können zur Optimierung der Anzahl und Auswahl von Messstellen angewandt werden, um unter Berücksichtigung der Kosten, räumlicher Abdeckung und Überwachungszielen die bestmögliche Auswahl an Messstellen zu treffen. Weiterhin kann die Eliminierung von Redundanzen in der Messstellenauswahl eine Rolle spielen. Auch kann die Optimierung (z. B. mit genetischen Algorithmen) mehrerer Kriterien gleichzeitig erfolgen. Weiterhin können Konzepte der Informationstheorie hilfreich sein. Diese Methoden fokussieren sich darauf, wie die Unsicherheit in einem System (Entropie) mittels neuer Informationen reduziert werden kann. Hierbei werden Standorte gewählt, die den größten Informationsgewinn liefern und beispielsweise die Unsicherheit im Gesamtsystem maximal reduzieren. Methoden wie die Entropieanalyse oder die Berechnung gegenseitiger Information können hierbei eine zentrale Rolle spielen. Nicht zuletzt ermöglichen die zuvor erwähnten numerischen Grundwassermodelle eine Optimierung der Messstellenanzahl und -platzierung innerhalb der Simulation.

7 Ausblick

Die Ergebnisse des Forschungsprojektes verdeutlichen, dass Wirkstoffe und die Abbauprodukte von Pflanzenschutzmitteln mit teils regional unterschiedlichen Belastungsmustern im Grundwasser nachgewiesen werden können. Der Einsatz maschineller Lernverfahren eröffnet dabei neue Möglichkeiten, um komplexe Zusammenhänge zwischen Standortbedingungen und dem Nachweis aus Grundwasser-Monitoringdaten besser zu verstehen und räumlich differenziert und stoffspezifisch zu bewerten. Interesse besteht darin, die im Projekt verwendeten ML-Modelle weiterzuentwickeln, um die Risikobewertung in der Zulassung von PSM im Rahmen der Verwendung von nationalen Nachzulassungsmonitoring (NZM)-Studien zu verbessern. Langfristig kann die Kombination aus Umweltmonitoring und künstlicher Intelligenz einen wichtigen Beitrag leisten, um die Qualität des Grundwassers zu sichern und den Einsatz von Pflanzenschutzmitteln nachhaltiger zu gestalten – im Einklang mit dem Vorsorgeprinzip und den Zielen des integrierten Pflanzenschutzes. Einige zentrale Anforderungen als kurzer Ausblick:

- ▶ **Datenbank:** Die Harmonisierung einer bundesweiten Datenbank ist ein erheblicher Aufwand und sollte nicht nur für PSM-Daten weitergeführt werden, sondern auch für Daten, denen ein großer Mehrwert für ML-Ansätze beigemessen wird. Ein Beispiel sind bundeslandspezifische lithostratigraphische Daten in der Form von Schichtverzeichnissen. Diese konnten im Rahmen des Projekts aufgrund des hohen Aufwands für die Harmonisierung nicht verwendet werden. Uneinheitliche Formate beziehen sich auf Stoffnamen von Wirkstoffen, wo teilweise unklar ist, ob es sich um denselben Stoff handelt sowie Stammdaten der Messstelle mit unterschiedlichen Bezeichnungen. Dem Potenzial zusätzlicher harmonisierter Features zum Trainieren von ML-Modellen durch bundesweite Daten des Grundwasserflurabstandes, Grundwasserfließrichtung und von Schichtenverzeichnissen wird ein großer Mehrwert beigemessen. Eine Verringerung und Harmonisierung der Bestimmungsgrenzen (LOQ) könnte dazu beitragen, bundesweite Wirkstofffunde zukünftig besser auswerten zu können.
- ▶ **Skala der Eingangsdaten:** Die Verfügbarkeit von Geodaten wächst stetig und doch sind verfügbare Geodaten häufig eine limitierende Größe für eine kohärente Bewertung der Grundwasser-Vulnerabilität. Eine Erhöhung der Maßstabszahl von 1:250.000 – wie in diesem Projekt anhand der hydrologischen und bodenkundlichen Eingangsdaten für das maschinelle Lernen – auf 1:1.000.000, wie es in Europa für EU-weit verfügbaren Geodaten diskutiert wird, ist für die Vorhersagegüte vieler Modelle als kritisch zu bewerten. Bodeneingangsdaten, die nur in wenigen Bundesländern in einer sehr hohen räumlichen Auflösung flächendeckend vorliegen (bspw. BK 1:5000), können einen Mehrwert bieten. Allerdings wurde der Einfluss des Skalenniveaus auf die Treffsicherheit einer Vorhersage weder in ML-Ansätzen noch für prozessorientierte *spatial distributed leaching models* (SDLM) systematisch untersucht.
- ▶ **Prozessverständnis von SDLMs:** Räumliche Modellansätze wurden bisher kaum anhand größerer Monitoringdatensätze validiert. Damit fehlt ein Abgleich, ob die empfohlene Risiko- bzw. Vulnerabilitätsbewertung durch SDLMs (bspw. GeoPELMO Modelle) die Realität hinreichend gut abbildet. Besonders validierungsbedürftig ist das mechanistische Prozessverständnis darüber, warum es an Standorten, die aufgrund ihrer Merkmale als gering gefährdet eingestuft wurden (0–25. Perzentil), dennoch zu einem Eintrag von PSM-

Wirkstoffen oder ihren Metaboliten ins Grundwasser kommt. Aktuell sind die Ursachen dafür spekulativ und die mechanistischen Zusammenhänge unklar.

- ▶ Aufwandmengen von PSM-Wirkstoffen: Aufwandmengen sind das entscheidende Bindeglied, um der Vulnerabilität eines Standortes eine Metrik geben zu können. Eine Indexentwicklung zur Abschätzung der intrinsischen Vulnerabilität ohne ackerschlagspezifische Kenntnis der Aufwandmenge ist nicht möglich. Die Bereitstellung dieser Daten in der Zukunft ist unabdingbar für ein mechanistisches Verständnis. Der wissenschaftliche Erkenntnisgewinn ist der entscheidende Mehrwert für die Praxis der Risikobewertung in der Zulassung und Regulatorik von PSM.

8 Quellenverzeichnis

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, P.; Warden, O.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. (2016), TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv. <https://doi.org/10.48550/arXiv.1603.04467>.

Ad-hoc-Arbeitsgruppe Boden (2005), Bodenkundliche Kartieranleitung: Mit 41 Abbildungen, 103 Tabellen und 31 Listen (5., verbesserte und erweiterte Auflage.). In Kommission: E. Schweizerbart'sche Verlagsbuchhandlung (Nägele und Obermiller). ed. Herausgegeben von der Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten der Bundesrepublik Deutschland (5. verbesserte und erweiterte Auflage; Stuttgart, Germany: Schweizerbart Science Publishers).

Aden, K.; Binner, R.; Fischer, R.; Gottschild, D.; Kloskowski, R.; Schinkel, K.; Michalski, B. (2002), Schutz des Grundwassers vor Pflanzenschutzmitteleinträgen: Leitlinie zur Aufklärung von Funden und zur Durchführung von zulassungsbegleitenden Monitoringstudien. – Nachrichtenbl. Deut. Pflanzenschutzd., 54 (5): 125-129.

Adlunger, K.; Banning, H.; Kuppe, K.; Osterwald, A.; Rauch, M.; Türkowsky, D. (2022), Nicht relevant? Abbauprodukte von Pflanzenschutzmitteln als Risiko für das Grundwasser.
<<https://www.umweltbundesamt.de/sites/default/files/medien/4031/publikationen/umid-onlineversion-01-2022.pdf>>(accessed 20.05.2025)

Albers, C. N. and Sültenfuss, J. (2024), A 60-Year Increase in the Ultrashort-Chain PFAS Trifluoroacetate and Its Suitability as a Tracer for Groundwater Age, *Environmental Science & Technology Letters*, 11 (10), 1090–95. <https://doi.org/10.1021/acs.estlett.4c00525>.

Albert, K. (2005), Remediation booklets 10. Commentary to the sensitivity maps. Környezetvédelmi és Vízügyi Minisztérium. Budapest http://www.kvvm.hu/szakmai/karmentes/kiadvanyok/eng/kf10.htm#fej_01.

Albinet, M. and Margat, J. (1970), Cartographie de la vulnérabilité à la pollution des nappes d'eau souterraines. Bull. BRGM, 2ème série, section 3 4: 13-22.

Aller, L.; Bennett, T.; Lehr, J.; Petty, R.; Hackett, G. (1987), DRASTIC: Standardized system for evaluating groundwater pollution potential using hydrogeologic settings, *Journal of the Geological Society of India*, 29. <https://doi.org/10.17491/jgsi/1987/290112>.

Amatulli, G.; Domisch, S.; Tuanmu, M.-N.; Parmentier, B.; Ranipeta, A.; Malczyk, J.; Jetz, W. (2018), A suite of global, cross-scale topographic variables for environmental and biodiversity modeling, *Scientific Data*, 5 (1), 180040. <https://doi.org/10.1038/sdata.2018.40>.

Arias-Estévez, M.; López-Periágo, E.; Martínez-Carballo, E.; Simal-Gándara, J.; Mejuto, J.-C.; García-Río, L. (2008), The mobility and degradation of pesticides in soils and the pollution of groundwater resources, *Agriculture, Ecosystems & Environment*, 123 (4), 247–60. <https://doi.org/10.1016/j.agee.2007.07.011>.

Bacci, E. and Gaggi, C. (1993), Simple models for ranking pesticide mobility from soils, in A. Del Re, et al. (eds.), *Proc. IX Symposium pesticide chemistry, mobility and degradation of xenobiotics*, 209–19.

Banning, H.; Bialek, K.; König, W.; Müller, A.; Pickl, C.; Scheithauer, M.; Straus, G.; Tüting, W. (2022), Empfehlungsliste für das Monitoring von Pflanzenschutzmittel-Metaboliten in deutschen Grundwässern. Umweltbundesamt (Hrsg.) Dessau.
https://www.umweltbundesamt.de/sites/default/files/medien/3521/dokumente/2022_07_29_uba_empfehlungsliste_update2022_de_0.pdf.

Bencini, A.; Cazzaroli, G.; Gargini, A.; Pranzini, G. (1995), La qualité des eaux souterraines et sa relation avec la vulnérabilité à la pollution des aquifères. Un exemple en Toscane (Italie) : la plaine de Florence." *Hydrogéologie* 3: 59-72.

BGR (2019), Mittlere jährliche Grundwasserneubildung von Deutschland 1:1.000.000 (GWN1000) (WMS). Bundesanstalt für Geowissenschaften und Rohstoffe (Hrsg) Hannover.

BGR (2023), Mittlere jährliche Sickerwasserrate aus dem Boden in Deutschland (SWR1000). Bundesanstalt für Geowissenschaften und Rohstoffe (Hrsg) Hannover.

BGR (2024), Bodenübersichtskarte der Bundesrepublik Deutschland 1:250.000 (BÜK250). Bundesanstalt für Geowissenschaften und Rohstoffe (Hrsg) Hannover, 10.25928/BUEK250_6.0.

BGR; LfU Bayern; LfU Brandenburg; LU Baden-Württemberg; HLNUG Hessen; LUNG Mecklenburg-Vorpommern; NLWKN Niedersachsen; LANUK Nordrhein-Westfalen; LfU Rheinland-Pfalz; LfU Schleswig-Holstein; LUA Saarland; LfULG Sachsen; LHW Sachsen-Anhalt; TLUBN Thüringen (2025), Bundesweiter Datensatz zu Pflanzenschutzmitteln im Grundwasser, <https://doi.org/10.25928/exxt-rh91>.

BGR und SGD (2005), Schutzpotenzial der Grundwasserüberdeckung. Bundesanstalt für Geowissenschaften und Rohstoffe (Hrsg) Hannover.

<https://www.bgr.bund.de/DE/Themen/Wasser/Projekte/abgeschlossen/Beratung/Sgwu/sgwu_projektbeschr.html>(accessed 23.04.2025)

BGR und SGD (2015), Hydrogeologische Raumgliederung von Deutschland (HYRAUM) (Engl. Trans.: Hydrogeological Spatial Structure of Germany (HYRAUM)), (Digitale Kartendaten v3.2. HYRAUM v32 edn.; Hannover, Germany.).

BGR und SGD (2016), Charakterisierung der Deckschichten gemäß LAWA-Arbeitshilfe - Methodische Übersicht. <https://www.bgr.bund.de/DE/Themen/Wasser/Projekte/abgeschlossen/Beratung/Sgwu/sgwu_methodenSGD.pdf;jsessionid=694C29354DE14B36729E4941B5FF4809.internet972?__blob=publicationFile&v=4>(accessed 24.04.2025)

BGR und SGD (2019), Hydrogeologische Übersichtskarte von Deutschland 1:250.000 (HÜK250), in Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) & Staatliche Geologische Dienste (SGD) (ed.), (Digitaler Datenbestand, Version 1.0.3 - Hannover).

BKG (2016), Digitales Basis-Landschaftsmodell (Ebenen) (Basis-DLM).

<<https://gdz.bkg.bund.de/index.php/default/digitales-basis-landschaftsmodell-kompakt-basis-dlm-kompakt.html>>(accessed 16.07.2025)

BKG (2018), Landbedeckungsmodell für Deutschland. <<https://gdk.gdi-de.org/geonetwork/srv/api/records/ED7F6DE9-6433-41DA-928A-044837C6EDAA>>(accessed 24.04.2025)

Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. (2022), Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany, *Remote Sensing of Environment*, 269. <https://doi.org/10.1016/j.rse.2021.112831>.

Bloomfield, J. P.; Williams, R. J.; Goody, D. C.; Cape, J. N.; Guha, P. (2006), Impacts of climate change on the fate and behaviour of pesticides in surface and groundwater—a UK perspective, *Science of The Total Environment*, 369 (1), 163–77. <https://doi.org/10.1016/j.scitotenv.2006.05.019>.

Breiman, L. (2001), Random Forests, *Machine Learning*, 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.

Carpenter, S. G. (1992), SEEPAGE: A system for early evaluation of the pollution potential of agricultural groundwater environments. USDA, Soil Conservation Service, Morgantown, WV, 23 p.

- Carter, A. D.; Palmer, R. C.; Monkhouse, R. A. (1987), Mapping the vulnerability of groundwater to pollution from agricultural practice, particularly with respect to nitrate. 38. International Conference, The Hague, The Netherlands.
- Chachadi, A. and Lobo-Ferreira, J. (2005), Assessing aquifer vulnerability to sea-water intrusion using GALDIT method: part 2 – GALDIT indicator descriptions. IAHS and LNEC, Proceedings of the 4th The Fourth Inter Celtic Colloquium on Hydrology and Management of Water Resources, held at Universidade do Minho, Guimarães, Portugal, July 11–13, 2005.
- Civita, M. and De Regibus, C. (1995), Sperimentazione di alcune metodologie per la valutazione della vulnerabilità degli acquiferi. Atii del 2o Convegno Nazionale sulla protezione e gestione delle acque sotterranee: metodologie, tecnologie e obiettivi, Nonantola (Modena).
- Civita, M. and Maio, M. (2008), Assessing and mapping groundwater vulnerability to contamination: The Italian "combined" approach, *Geofísica internacional*, ISSN 0016-7169, Vol. 43, Nº. 4, 2004, pags. 513-532, 43. 10.22201/igeof.00167169p.2004.43.4.776.
- Connor, J. T.; Martin, R. D.; Atlas, L. E. (1994), Recurrent neural networks and robust time series prediction, *IEEE Transactions on Neural Networks*, 5 (2), 240–54. <https://dx.doi.org/10.1109/72.279188>.
- Cooke, A.-K.; Willkommen, S.; Broda, S. (2024), Analysing agricultural plant protection product concentrations in groundwater in Germany: Nationwide database with site and compound insights, *Environmental Research*, 248, 118231. <https://doi.org/10.1016/j.envres.2024.118231>.
- Cooke, A.-K.; Mohanty, S.; Joger, F.; Erkens, E.; Broda, S. (2025), Spatial modelling of plant protection product metabolites in groundwater in Germany with Random Forest and Feedforward Neural Networks. In preparation for submission to *Hydrology and Earth System Sciences*.
- Cramér, H. (1946), The two-dimensional case, in H. Cramér (ed.), *Mathematical Methods of Statistics* (Princeton: Princeton University Press), 282.
- Daly, D. and Warren, W. P. (1998), Mapping groundwater vulnerability: the Irish perspective. London, Geological Society, Special Publications. <https://doi.org/10.1144/GSL.SP.1998.130.01.16>.
- Das, D.; Avancha, S.; Mudigere, D.; Vaidynathan, K.; Sridharan, S.; Kalamkar, D.; Kaul, B.; Dubey, P. (2016), Distributed Deep Learning Using Synchronous Stochastic Gradient Descent, arXiv. <https://doi.org/10.48550/arXiv.1602.06709>.
- Davis, M.; Coker, G.; Watt, M.; Graham, D.; Pearce, S.; Dando, J. (2012), Nitrogen leaching after fertilising young *Pinus radiata* plantations in New Zealand, *Forest Ecology and Management*, 280, 20-30. <https://doi.org/10.1016/j.foreco.2012.06.009>.
- DELG, E., GSI (1999), Groundwater Protection Schemes. Department of the Environment and Local Government, Environmental Protection Agency and Geological Survey of Ireland, 24 pp. 07.07.2025)
- Deutsches Institut für Normung (1997), DIN 19732: Bestimmung des standörtlichen Verlagerungspotentials von nichtsorbiebaren Stoffen.- 4 S.; Berlin.
- Doerfliger, N.; Jeannin, P. Y.; Zwahlen, F. (1999), Water vulnerability assessment in karst environments: a new method of defining protection areas using a multi-attribute approach and GIS tools (EPIK method), *Environmental Geology*, (39), 165–76. <https://doi.org/10.1007/s002540050446>.
- DWD (2025), Klimawandel – ein Überblick
<https://www.dwd.de/DE/klimaumwelt/klimawandel/klimawandel_node.html>(accessed 24.04.2025)
- EFSA PPR Panel; Hernandez-Jerez, A.; Adriaanse, P.; Aldrich, A.; Berny, P.; Coja, T.; Duquesne, S.; Focks, A.; Marinovich, M.; Millet, M.; Pelkonen, O.; Pieper, S.; Topping, C.; Widenfalk, A.; Wilks, M.; Wolterink, G.; Kasteel, R.; Kuppe, K.; Tiktak, A. (2023), Statement of the Scientific Panel on Plant Protection Products and their

Residues (PPR Panel) on the design and conduct of groundwater monitoring studies supporting groundwater exposure assessments of pesticides, 21 (5), e07990. <https://doi.org/10.2903/j.efsa.2023.7990>.

EU (2000), Water Framework Directive - EU Directive 2000/60/EC of 22 December 2000 establishing a framework for community action in the field of water policy, (327: European Union), 1-72.

EU (2006), Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on protection of groundwater against pollution and deterioration, OJ L 372, 27.12. 2006, 19–31, (European Union).

EU (2024), Verwaltung der Zahlungen - Europäische Kommission (europa.eu).

<https://agriculture.ec.europa.eu/common-agricultural-policy/financing-cap/assurance-and-audit/managing-payments_de>(accessed 02.09.2024)

FAO (2023), Pesticides use and trade (1990–2021), FAOSTAT ANALYTICAL BRIEF 70.

<<https://openknowledge.fao.org/items/fa226b62-88ab-4637-bf34-445361d6c312>>(accessed 10.04.2025)

Fisher, A.; Rudin C.; F., D. (2019), All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously, arXiv. <https://arxiv.org/abs/1801.01489>.

Fitzsimons, V.; Daly, D.; Deakin, J. (2003), GIS guidelines for the assessment and mapping of groundwater vulnerability to contamination. Draft June 2003. 79 p.

Fobe, B. and Goossens, M. (1990), The groundwater vulnerability map for the Flemish region: its principles and uses. , *Engineering Geology*, (29), 355–63. [https://doi.org/10.1016/0013-7952\(90\)90069-D](https://doi.org/10.1016/0013-7952(90)90069-D).

FOCUS (2009/2014), Assessing Potential for Movement of Active Substances and their Metabolites to Ground Water in the EU – The Final Report of the Ground Water Work Group of FOCUS. Updated by EFSA at the request of the European Commission to incorporate pertinent aspects of the EFSA PPR panel opinions in version 1 of 13 June 2009 – Sanco/13144/2010, version 3, 10 October 2014.

Foster, S. S. D. (1987), Fundamental concepts in aquifer vulnerability pollution risk and protection strategy. Proceedings of International Conference: Vulnerability of Soil and Groundwater to Pollutants, Noordwijk, The Netherlands.

Fundaufklärungsverfahren, N. E. (2016), Straffung und Effektivierung der Fundaufklärung
Maßnahmenvorschlag der UAG „Trinkwasser“ zum Beschlussvorschlag der AG „Pflanzenschutz und Gewässerschutz“ vom 25.11.2016 für das NAP-Forum am 1./2.12.2016. <https://www.nap-pflanzenschutz.de/fileadmin/SITE_MASTER/content/Bilder/Forum_NAP/Forum_Dez_2016/161125_Forum_NA P_Fundaufklaerung_Hintergrundpapier.pdf>(accessed

Gareth, J.; Witten, D.; Hastie, T.; Tibshirani, R. (2023), An introduction to statistical learning with Applications in R. (Second Ed.; New York: Springer).

Gimsing, A. L.; Agert, J.; Baran, N.; Boivin, A.; Ferrari, F.; Gibson, R.; Hammond, L.; Hegler, F.; Jones, R. L.; König, W.; Kreuger, J.; van der Linden, T.; Liss, D.; Loiseau, L.; Massey, A.; Miles, B.; Monrozier, L.; Newcombe, A.; Poot, A.; Reeves, G. L.; Reichenberger, S.; Rosenbom, A. E.; Staudenmaier, H.; Sur, R.; Schwen, A.; Stemmer, M.; Tüting, W.; Ulrich, U. (2019), Conducting groundwater monitoring studies in Europe for pesticide active substances and their metabolites in the context of Regulation (EC) 1107/2009, *Journal of Consumer Protection and Food Safety*, 14 (1), 1-93. <https://doi.org/10.1007/s00003-019-01211-x>.

Goldscheider, N.; Klute, M.; Sturm, S.; Hötzl, H. (2000), The PI method – a GIS-based approach to mapping groundwater vulnerability with special consideration of karst aquifers. *Zeitschrift für angewandte Geologie* 46(3): 157-166.

Goodfellow, I.; Bengio, Y.; Courville, A. (2016), *Deep Learning* (MIT Press).

- Goyal, P.; Dollár, P.; Girshick, R.; Noordhuis, P.; Wesolowski, L.; Kyrola, A.; Tulloch, A.; Jia, Y.; He, K. (2018), Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour, arXiv. <https://doi.org/10.48550/arXiv.1706.02677>.
- Griffiths, P.; Nendel, C.; Hostert, P. (2019), Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping, *Remote Sensing of Environment*, 220, 135-51. <https://doi.org/10.1016/j.rse.2018.10.031>.
- GrwV (2022), Grundwasserverordnung vom 9. November 2010 (BGBl. I S. 1513), die zuletzt durch Artikel 1 der Verordnung vom 12. Oktober 2022 (BGBl. I S. 1802) geändert worden ist.
- Haertle, T. and Josopait, V. (1982), Methodik und Arbeitsweise zur Anfertigung von Karten über die natürlichen Grundwasserschutzbedingungen. Inst. Stadtbauwesen TU Braunschweig (Hrsg.): Anthropogene Einflüsse auf die Grundwasserbeschaffenheit in Niedersachsen. Band 34, 91-110.
- Halbach, K.; Möder, M.; Schrader, S.; Liebmann, L.; Schäfer, R. B.; Schneeweiss, A.; Schreiner, V. C.; Vormeier, P.; Weisner, O.; Liess, M.; Reemtsma, T. (2021), Small streams—large concentrations? Pesticide monitoring in small agricultural streams in Germany during dry weather and rainfall, *Water Research*, 203, 117535. <https://doi.org/10.1016/j.watres.2021.117535>.
- Heinkele, T.; Jahnke, C.; Voigt, H. J.; Hannappel, S.; Donat, E. (2002), Charakterisierung der Empfindlichkeit von Grundwasserkörpern. Umweltbundesamt (Hrsg.) Dessau. <https://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/2132.pdf>.
- Helsel, R. D. (2012), *Statistics for censored environmental data using minitab and R* (Second edn., Wiley series in statistics in practice; Hoboken, New Jersey: John Wiley & Sons).
- Höltling, B.; Haertlé, T.; Hohberger, K. H.; Nachtigall, K. H.; Villinger, E.; Weinzierl, W.; Wrobel, J. P. (1995), Konzept zur Ermittlung der Schutzfunktion der Grundwasserüberdeckung, *Geologisches Jahrbuch C*, 63, 5-24.
- Kavouri, K.; Plagnes, V.; Tremoulet, J.; Dörfliker, N.; Rejiba, F.; Marchet, R. (2011), PaPRIKa: a method for estimating karst resource and source vulnerability - application to the Ouyse karst system (southwest France). *Hydrogeology Journal* 19(2): 339-353.
- Keskar, N.; Mudigere, D.; Nocedal, J.; Smelyanskiy, M.; Tang, P. (2017), On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima, arXiv. <https://doi.org/10.48550/arXiv.1609.04836>.
- Kingma, D. P. and Ba, J. (2015), Adam: A Method for Stochastic Optimization, arXiv. <https://doi.org/10.48550/arXiv.1412.6980>.
- Klein, M.; Thomas, K.; Trapp, M.; Guerniche, D. (2019), Protection of the groundwater against loads of plant protection products: validation of the new EU-simulation model FOCUS PELMO 4 for a reliable prediction of the leaching potential of PPP into groundwater. Part B: GIS-based analysis of the protection level of the FOCUS-scenarios representative for Germany concerning climate and soil properties for the national risk assessment groundwater. – UBA Texte 146/2019. https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-12-05_texte_146-2019_focus-pelmo-4_part-b.pdf (accessed 16.07.2025)
- Klein, M.; Thomas, K.; Trapp, M.; Hannappel, S.; Seifert, S. (2025), Protect groundwater from plant protection products. Spatial distributed leaching modelling to identify agricultural areas with high risk for leaching. Abschlussbericht Refoplan-Projekt FKZ 3716674090. Im Veröffentlichungsprozess.
- Knoll, L.; Breuer, L.; Bach, M. (2019), Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning, *Science of The Total Environment*, 668, 1317–27. <https://doi.org/10.1016/j.scitotenv.2019.03.045>.

König, W.; Bach, T.; Börke, P.; Burucker, G.; Feuerstein, W.; Gathmann, A.; Haenel, S.; Handke, H.; Hilliges, F.; Jankowski, A.; Marahrens, S.; Müller, A.; Pfannerstill, M.; Pickl, C.; Rauch, M.; Reuther, C.; Simon-O'Malley, S.; Straus, G.; Tüting, W.; Wieger, C. (2020), Nachzulassungsmonitoring im Zulassungsverfahren von Pflanzenschutzmitteln : Messstellen Vorauswahl durch die Bundesländer. Umweltbundesamt (Hrsg.) Dessau. https://www.umweltbundesamt.de/nzm_messstellenauswahl.

https://www.umweltbundesamt.de/sites/default/files/medien/362/dokumente/messstellenauswahl_nzm-studien_oct2020.pdf.

Lang et al. (2019), mlr3: A modern object-oriented machine learning framework in R, *Journal of Open Source Software*, 4, 1903. <https://doi.org/10.21105/joss.01903>.

Larsen, A. E.; Quandt, A.; Foxfoot, I.; Parker, N.; Sousa, D. (2023), The effect of agricultural land retirement on pesticide use, *Science of The Total Environment*, 896, 165224.

<https://doi.org/10.1016/j.scitotenv.2023.165224>.

LAWA (2003a), Arbeitsexemplar LAWA-Arbeitshilfe zur Umsetzung der WRRL / Teil 3.

<https://www.bgr.bund.de/DE/Themen/Wasser/Projekte/abgeschlossen/Beratung/Sgwu/lawa_arbeitshilfe.pdf;jsessionid=694C29354DE14B36729E4941B5FF4809.internet972?__blob=publicationFile&v=4>(accessed 24.04.2025)

LAWA (2003b), Arbeitshilfe zur Umsetzung der EG Wasserrahmenrichtlinie, Teil 3, 36–37: Charakterisierung der Deckschichten.

<https://www.bgr.bund.de/DE/Themen/Wasser/Projekte/abgeschlossen/Beratung/Sgwu/lawa_arbeitshilfe.pdf?__blob=publicationFile&v=4>(accessed 07.07.2025)

LAWA (2018), Ermittlung von Verweil- und Fließzeiten. Abschlussbericht Arbeitshilfe zur Umsetzung der EG-Wasserrahmenrichtlinie. Ermittlung von Verweilzeiten in der ungesättigten und gesättigten Zone. 91 S.

<https://www.laenderfinanzierungsprogramm.de/static/LFP/Dateien/LAWA/AG/G_1.17_Anlage_1_Kurzanleitung_zur_Ermittlung_von%20Flie%C3%9F-_und_%20Reaktionszeiten_GWK-Messnetz.pdf>(accessed 07.07.2025)

LAWA (2024), Bericht zur Grundwasserbeschaffenheit in Deutschland. Pflanzenschutzmittelwirkstoffe und Metaboliten. Funde und Tendenzen. Berichtszeitraum 2017 bis 2021. <https://www.lawa.de/documents/psm-bericht-2023-12-22-barrierearm-final_2_1728974845.pdf>(accessed 07.07.2025)

Lewis, K. A.; Tzilivakis, J.; Warner, D.; Green, A. (2016), An international database for pesticide risk assessments and management, *Human and Ecological Risk Assessment: An International Journal*, 22 (4), 1050–64.

<https://doi.org/10.1080/10807039.2015.1133242>.

Liess, M.; Böhme, A.; Gröning, J.; Liebmann, L.; Lück, M.; Reemtsma, T.; Römerscheid, M.; Schade, U.; Schwarz, B.; Vormeier, P.; Weisner, O. (2019), Belastung von kleinen Gewässern in der Agrarlandschaft mit Pflanzenschutzmittel-Rückständen – TV1 Datenanalyse zur Pilotstudie Kleingewässermonitoring 2018/2019.

Umweltbundesamt (Hrsg.) Dessau. , Abschlussbericht, 63/2023.

https://www.umweltbundesamt.de/sites/default/files/medien/11850/publikationen/63_2023_texte_belastung_von_kleinen_gewaessern_in_der_agrarlandschaft_mit_pflanzenschutzmittel-rueckstaenden.pdf.

Luo, Y.; Atashgahi, S.; Rijnaarts, H. H. M.; Comans, R. N. J.; Sutton, N. B. (2019), Influence of different redox conditions and dissolved organic matter on pesticide biodegradation in simulated groundwater systems, *Science of The Total Environment*, 677, 692–99. <https://doi.org/10.1016/j.scitotenv.2019.04.128>.

Mádl-Szonyi, J. and Füle, L. (1998), Groundwater vulnerability assessment of the SW Trans-Danubian Central Range, Hungary. *Environmental Geology* 35(1): 9–18. <https://doi.org/10.1007/s002540050287>.

Magiera, P. (2000), Methoden zur Abschätzung der Verschmutzungsempfindlichkeit des Grundwassers. , *Grundwasser*, 3, 103–14. <https://doi.org/10.1007/s767-000-8357-1>.

- Malik, P.; Fendek, M.; Vrana, K.; Witkowski, A. (1998), Groundwater vulnerability map of the Muranska Planina Plateau. Conference: "Gambling with groundwater - Physical, Chemical, and Biological Aspects of Aquifer-Stream Relations" At: Las Vegas, USA Volume: "Gambling with groundwater - Physical, Chemical, and Biological Aspects of Aquifer-Stream Relations", American Institute of Hydrology, Paul, USA, p. 623-628.
- Malley, J. D.; Kruppa, J.; Dasgupta, A.; Malley, K. G.; Ziegler, A. (2012), Probability Machines, Consistent Probability Estimation Using Nonparametric Learning Machines, 51 (01), 74-81. [10.3414/ME00-01-0052](https://doi.org/10.3414/ME00-01-0052).
- Manamsa, K.; Crane, E.; Stuart, M.; Talbot, J.; Lapworth, D.; Hart, A. (2016), A national-scale assessment of micro-organic contaminants in groundwater of England and Wales, Science of The Total Environment, 568, 712-26. <https://doi.org/10.1016/j.scitotenv.2016.03.017>.
- Mardhel, V. and Gravier, A. (2006), Map of the vulnerability simplified Groundwater in the Loire-Brittany Basin (BRGM / RP-54553-FR)
- Margat, J. (1968), Vulnérabilité des nappes d'eau souterraine à la pollution.- BRGM-Veröffentlichung 68 SGL 198 HYD; Orléans.
- Maxe, L. and Johansson, P. O. (1998), Assessing groundwater vulnerability using travel time and specific surface area as indicators. Hydrogeology Journal 6: 441-449. <https://doi.org/10.1007/s100400050166>.
- McRae, B. (1989), The characterization and identification of potentially leachable pesticides and areas vulnerable to groundwater contamination by pesticides in Canada. Backgrounder 89-01. A. C. Pesticides Directorate. Ottawa, Ontario.
- Meinardi, C. R.; Beusen, A. H. W.; Bollen, M. J. S.; Klepper, O.; Willems, W. J. (1995), Vulnerability to diffuse pollution and average nitrate contamination of European soils and groundwater. Water Science and Technology 31(8): 159-165. <https://doi.org/10.2166/wst.1995.0288>.
- Mohaupt, V.; Völker, J.; Altenburger, R.; Birk, S.; Kirst, I.; Kühnel, D.; Küster, E.; Semeradova, S.; Šubelj, G.; Whalley, C. (2020), Pesticides in European rivers, lakes and groundwaters - Data assessment, (European Topic Centre on Inland, Coastal and Marine waters (ETC/ICM)).
- Molnar, C. (2022), Interpretable Machine Learning - A Guide for Making Black Box Models Explainable (2nd Edition edn.: Leanpub).
- Nölscher, M.; Mutz, M.; Broda, S. (2022), Multiorder hydrologic Position for Europe — a Set of Features for Machine Learning and Analysis in Hydrology, Scientific Data, 9 (1), 662. <https://doi.org/10.1038/s41597-022-01787-4>.
- Ohmer, M.; Liesch, T.; Goldscheider, N. (2019), On the Optimal Spatial Design for Groundwater Level Monitoring Networks, 55 (11), 9454-73. <https://doi.org/10.1029/2019WR025728>.
- Ohmer, M.; Liesch, T.; Wunsch, A. (2022), Spatiotemporal optimization of groundwater monitoring networks using data-driven sparse sensing methods, Hydrol. Earth Syst. Sci., 26 (15), 4033-53. <https://doi.org/10.5194/hess-26-4033-2022>.
- Palmer, R. C. and Lewis, M. A. (1998), Assessment of groundwater vulnerability in England and Wales. London, Geological Society, Special Publications. <https://doi.org/10.1144/GSL.SP.1998.130.01.17>.
- PAPA (2024), Statistische Erhebung zur Anwendung von Pflanzenschutzmitteln in der Praxis (PAPA). <https://papa.julius-kuehn.de/> (accessed 29.10.2024)
- Pérez-Lucas, G.; Navarro, G.; Navarro, S. (2024), Adapting agriculture and pesticide use in Mediterranean regions under climate change scenarios: A comprehensive review, European Journal of Agronomy, 161, 127337. <https://doi.org/10.1016/j.eja.2024.127337>.
- Polyak, B. T. (1964), Some methods of speeding up the convergence of iteration methods, USSR Computational Mathematics and Mathematical Physics, 4 (5), 1-17. [https://doi.org/10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5).

- Preidl, S.; Lange, M.; Doktor, D. (2020), Introducing APiC for regionalised land cover mapping on the national scale using Sentinel-2A imagery, *Remote Sensing of Environment*, 240, 111673. <https://doi.org/10.1016/j.rse.2020.111673>.
- Probst, P.; Wright, M. N.; Boulesteix, A.-L. (2019), Hyperparameters and tuning strategies for random forest, 9 (3), e1301. <https://doi.org/10.1002/widm.1301>.
- Rao, P.; Hornsby, A.; Jessup, R. (1985), Indices for ranking the potential for pesticide contamination of groundwater, *Soil and Crop Science Society of Florida Proceedings*, 44, 1–8.
- Ravbar, N. and Godscheider, N. (2007), Proposed methodology of vulnerability and contamination risk mapping for the protection of Karst aquifers in Slovenia. *Acta Carsologica* 36/3, 397-411. <https://doi.org/10.3986/ac.v36i3.174>
- Ray, J. A. and O'dell, P. W. (1993), DIVERSITY: A new method for evaluating sensitivity of groundwater to contamination. *Environmental Geology* 22: 345-352. <https://doi.org/10.1007/BF00767508>.
- Rine, J. M.; Berg, R. C.; Shafer, J. M.; Covington, E. R.; Reed, J. K.; Bennett, C. B.; Trudnak, J. E. (1998), Development and testing of a contamination potential mapping system for a portion of the General Separations Area, Savannah River Site, South Carolina. *Environmental Geology* 35(4): 263-277. <https://doi.org/10.1007/s002540050313>.
- Roeper, U. V. R. (1990), Regina aquifers sensitivity mapping and land use guidelines. Report WQ 134. S. E. a. P. S. Water Quality Branch. Regina, SK.
- Rosenbom, A. E.; Olsen, P.; Plauborg, F.; Grant, R.; Juhler, R. K.; Brüsch, W.; Kjær, J. (2015), Pesticide leaching through sandy and loamy fields – Long-term lessons learnt from the Danish Pesticide Leaching Assessment Programme, *Environmental Pollution*, 201, 75-90. <https://doi.org/10.1016/j.envpol.2015.03.002>.
- Scharp, C.; Alveteg, T.; Johansson, P. O. (1997), Assigning a groundwater protection value: methodology development. Rotterdam, Balkema.
- Schaul, T.; Antonoglou, I.; Silver, D. (2014), Unit Tests for Stochastic Optimization, arXiv. <https://doi.org/10.48550/arXiv.1312.6055>.
- Schreinemachers, P. and Tipraqsa, P. (2012), Agricultural pesticides and land use intensification in high, middle and low income countries, *Food Policy*, 37 (6), 616-26. <https://doi.org/10.1016/j.foodpol.2012.06.003>.
- Seelig, B. (1994), An assessment system for potential groundwater contamination from agricultural pesticide use in North Dakota. *Extension Bulletin*. 63. <<https://library.ndsu.edu/server/api/core/bitstreams/5f4f4464-d3eb-485a-bc13-6109083b4461/content>>(accessed 07.07.2025)
- Sheng, G.; Johnston, C. T.; Teppen, B. J.; Boyd, S. A. (2001), Potential Contributions of Smectite Clays and Organic Matter to Pesticide Retention in Soils, *Journal of Agricultural and Food Chemistry*, 49 (6), 2899-907. <https://doi.org/10.1021/jf001485d>.
- Sheng, G.; Yang, Y.; Huang, M.; Yang, K. (2005), Influence of pH on pesticide sorption by soil containing wheat residue-derived char, *Environmental Pollution*, 134 (3), 457-63. <https://doi.org/10.1016/j.envpol.2004.09.009>.
- Sokol, G.; Leibundgut, C.; Schulz, K. P.; Weinzierl, W. (1993), Mapping procedures for assessing groundwater vulnerability to nitrate and pesticides. *HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources*, Vienna.
- Spearman, C. (1904), The Proof and Measurement of Association between Two Things, *The American Journal of Psychology*, 15 (1), 72-101. <https://doi.org/10.2307/1412159>.
- Stempvoort, D. V.; Ewert, L.; Wassenaar, L. (1993), Aquifer vulnerability index: a GIS-compatible method for groundwater vulnerability mapping, *Canadian water resources journal*, 18 (1), 25-37. <https://doi.org/10.4296/cwrj1801025>.

- Tiktak, A.; Poot, A.; Jene, B.; Ghafoor, A.; Berg, E. v. d.; Hoogeweg, G.; Klein, M.; Stemmer, M.; Sur, R.; Sweeney, P. (2020), Spatially Distributed Leaching Modelling of Pesticides in the context of Regulation (EC) 1107/2009 - Problem definition document. <https://esdac.jrc.ec.europa.eu/public_path/shared_folder/SDLM/SETAC-SDLM-problem-definition-document-2020.pdf>(accessed 07.07.2025)
- Tiktak, A.; Poot, A.; Jene, B.; Hoogeweg, G.; Braakhekke, M.; Wipfler, L.; Cornelissen, P.; Klein, J.; Stemmer, M.; Ritter, A.; Sur, R.; Spickermann, G.; Heuvelink, G.; Hughes, G.; Marahrens, S.; Reichenberger, S.; Suci, N.; Morris, M. (2025), Harmonised Framework for the SETAC Spatially Distributed Leaching Modelling of Pesticides Initiative: 2025 Update in SETAC Europe 35Th Annual Meeting, 3.03.A.T.-02, Abstract Book, SETAC Europe 35th Annual Meeting, Vienna, Austria, S. 407.
<<https://setac.confex.com/setac/europe2025/meetingapp.cgi/Paper/29302>>(accessed 16.07.2025)
- Van Stempvoort, D.; Ewert, L.; Wassenaar, L. (1993), Aquifer vulnerability index: a GIS-compatible method for groundwater vulnerability mapping." *Canadian Water Resources Journal* 18(1): 25-37.
- Vias, J. M.; Andreo, B.; Perles, M. J.; Carrasco, F. (2006), Proposed method for the groundwater vulnerability mapping in carbonate (karstic) aquifers: the COP method. Application in two pilot sites in Southern Spain, *Hydrogeology Journal*, 14, 912-25. <https://doi.org/10.1007/s10040-006-0023-6>.
- Vinten, A. J. A.; Vivian, B. J.; Wright, F.; Howard, R. S. (1994), A comparative study of nitrate leaching from soils of differing textures under similar climatic and cropping conditions, *Journal of Hydrology*, 159 (1), 197-213. [https://doi.org/10.1016/0022-1694\(94\)90256-9](https://doi.org/10.1016/0022-1694(94)90256-9).
- Voigt, H. J. (1987), Nutzung und Schutz der Gewässer, Grundwässer, Klassifizierung. TGL 34 334. Halle.
<https://katalog.ub.uni-weimar.de/tgl/TGL_34334_05-1986.pdf>(accessed 07.07.2025)
- Weil, R. and Brady, N. (2017), *The Nature and Properties of Soils*. 15th edition.
- Whitfield, M. (1974), Thermodynamic limitations on the use of the platinum electrode in Eh measurements, 19 (5), 857-65. <https://doi.org/10.4319/lo.1974.19.5.0857>.
- Wirsing, T.; Neukum, C.; Goldscheider, N.; Maier, M. (2015), Integration der bodenkundlichen Filter- und Pufferfunktion in die hydrogeologische Vulnerabilitätsbewertung, *Grundwasser*, 20 (2), 97-106. <https://doi.org/10.1007/s00767-014-0273-5>.
- Wright, M. N. and Ziegler, A. (2017), ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, *Journal of Statistical Software*, 77 (1), 1 - 17. <https://doi.org/10.18637/jss.v077.i01>.
- Wunsch, A.; Liesch, T.; Broda, S. (2022), Deep learning shows declining groundwater levels in Germany until 2100 due to climate change, *Nature Communications*, 13 (1), 1221. <https://doi.org/10.1038/s41467-022-28770-2>.
- You, Y.; Gitman, I.; Ginsburg, B. (2017), Large Batch Training of Convolutional Networks, arXiv. <https://doi.org/10.48550/arXiv.1708.03888>.
- Zelykman, E. M.; Lushchyk, A. V.; Eremenko, H. K. (1994), Quantitative assessment and the forecast of the protective ability of the upper part of the geological environment (soils of the unsaturated zone) [Kolychestvennaia otsenka y prohnaz zashchytnoi sposobnosti verkhnnei chasty heolohycheskoi sredy (hruntov zonu aeratsyy)]. Simferopol, 271 p., [in Russian].
- Zhang, Y.; Pinder, G. F.; Herrera, G. S. (2005), Least cost design of groundwater quality monitoring networks, *Water Resources Research*, 41 (8). <https://doi.org/10.1029/2005WR003936>.

A Anhang

A.1 Disclaimer zur Verwendung der Daten des FARM-Projektes und Informationen zur Datennachnutzung

Im Rahmen dieses Projekts wurden erstmalig Monitoringdaten aller Flächenbundesländer für eine Vielzahl an Stoffen, die an rund 26.000 Messstellen erhoben wurden, in einer zentralen Datenbank zusammengeführt, harmonisiert und ausgewertet. Das ursprüngliche Stoffspektrum umfasste Konzentrationsinformationen zu insgesamt 530 Pflanzenschutzmittel-Wirkstoffen und ihren Metaboliten im Grundwasser, alle diese Substanzen sind im Gesamtdatensatz enthalten. Für die Analyse wurde auf Basis der Informationsdichte (u. a. Flächenrepräsentativität, Anteil linkszensierter Daten) eine gezielte Vorauswahl getroffen, wodurch das Stoffspektrum auf 266 Substanzen reduziert wurde. Die darauf basierenden Ergebnisse des maschinellen Lernens sowie die Vulnerabilitätsbewertung des Grundwassers beziehen sich auf Daten von insgesamt 21.041 Messstellen.

Für diese Messstellenanzahl wurden verschiedene Eingangsdaten (features) für das maschinelle Lernen verwendet, die entweder über das Geoportal der BGR oder über spezifische Datenportale der Länder, teilweise frei verfügbar, bezogen wurden.

Die Datenübermittlung durch die Landesbehörden erfolgte zur wissenschaftlichen Nutzung und umfasste Informationen zu Messstellen des jeweiligen Landesmessnetzes sowie vereinzelt Daten von Wasserversorgern oder Dritten. Aufgrund der teils sensiblen Informationen mit Bezug auf konkrete Messstellen und Brunnen bestehen Einschränkungen hinsichtlich der Veröffentlichung und Nachnutzung des im FARM-Projekt harmonisierten Datensatzes.

Ein Teil der im Projekt verwendeten Messstellen konnte aufgrund von Vorgaben zu Besitzverhältnissen – insbesondere bei Daten von Wasserversorgern oder Dritten – nicht in den veröffentlichten Datensatz zur Nachnutzung aufgenommen werden. Diese Messstellen wurden vollständig entfernt (4,6% des Gesamtdatensatzes, 962 Messstellen, Tabelle 7).

Bei den verbleibenden Messstellen galten teils besondere Anforderungen an die Anonymisierung. Dazu zählten etwa die Veröffentlichung gerundeter statt genauer Koordinaten sowie die Anonymisierung der Messstellenkennungen (IDs). Es durften weiterhin keine Angaben zu Besitzverhältnissen gemacht werden.

Bitte beachten Sie, dass bei einer Weitergabe oder Veröffentlichung der Daten die jeweiligen länderspezifischen Nutzungsbedingungen einzuhalten sind. Diese sind auf den Geoportalen der unten genannten Landesämter einsehbar. Bei der Teilauswahl der Daten ist auf die Nennung der bundeslandspezifischen Quelle hinzuweisen. Die Quellenangabe muss eindeutig sein und ist für private, nicht kommerzielle sowie für wissenschaftliche Zwecke vorgesehen.

Die folgenden Flächenbundesländer haben dankenswerterweise Daten bereitgestellt:

- ▶ Brandenburg (Landesamt für Umwelt Brandenburg)
- ▶ Baden-Württemberg (Landesanstalt für Umwelt Baden-Württemberg)
- ▶ Bayern (Bayerisches Landesamt für Umwelt)
- ▶ Hessen (Hessisches Landesamt für Naturschutz, Umwelt und Geologie)
- ▶ Mecklenburg-Vorpommern (Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern)

- ▶ Niedersachsen (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz)
- ▶ Nordrhein-Westfalen (Landesamt für Natur, Umwelt und Klima Nordrhein-Westfalen)
- ▶ Rheinland-Pfalz (Landesamt für Umwelt Rheinland-Pfalz)
- ▶ Schleswig-Holstein (Landesamt für Umwelt des Landes Schleswig-Holstein)
- ▶ Saarland (Landesamt für Umwelt und Arbeitsschutz Saarland)
- ▶ Sachsen (Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie)
- ▶ Sachsen-Anhalt (Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt)
- ▶ Thüringen (Thüringer Landesamt für Umwelt, Bergbau und Naturschutz)

Tabelle 7: Zusammenfassung der Messstellen die für die Vulnerabilitätsbewertung des Grundwassers durch Pflanzenschutzmittel im FARM-Projekt durch die Länder bereitgestellt und ausgewertet wurden und Anzahl der Messstellen die für die Datennachnutzung veröffentlicht werden dürfen.

Bundesland	MST ausgewertet (n)	MST veröffentlicht (n)	Anteil gelöschter MST (%)
BB	808	808	0
BW	2363	2363	0
BY	2403	1460	39,2
HE	4025	4024	0
MV	657	657	0
NI	6005	6005	0
NRW	1639	1639	0
RP	716	716	0
SH	347	347	0
SL	54	36	33,3
SN	1227	1227	0
ST	563	563	0
TH	234	234	0
Gesamt	21.041	20.079	4,6

Der finalisierte und harmonisierte Datensatz ist auf dem Geoportal der BGR hinterlegt (BGR et al. 2025).

A.2 Informationen zu den Eingangsdaten (features) die für das maschinelle Lernen verwendet wurden.

Tabelle 8: Übersicht zu der reduzierten und erweiterten Auswahl der Eingangsdaten (features) die für das maschinelle Lernen verwendet wurden.

Konfiguration	Feature	Einheit	Erläuterung
Reduziert	Flächen pro Kulturklasse, für die der jeweilige Wirkstoff zugelassen ist	(%)	Aus Sentineldaten abgeleitete Kulturklasse im Puffer der Grundwassermessstelle. (Blickensdörfer et al. 2022)
Reduziert	Filteroberkante	(Meter unter Geländeoberkante)	Tiefe der Filteroberkante unter Geländeoberkante entnommen aus den Stammdaten der Grundwassermessstelle.
Reduziert	Sandgehalt im Boden	(%)	Tiefengewichteter Anteil des Sandgehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
Reduziert	Organischer Kohlenstoffanteil im Boden	(%)	Tiefengewichteter Anteil an organischen Kohlenstoff abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
Reduziert	Makroporenklasse vom Boden	(-)	Differenzierung der Klassen: no macropore flow (1), moderate macropore flow (2), high macropore flow (3); (Klein et al. 2025)
Reduziert	Langjährige Rate der Grundwasserneubildung	(mm/Jahr)	Mittlere langjährige Rate der Grundwasserneubildung (BGR 2019)
Reduziert	Hohlraumart des Aquifers	(-)	Differenzierung der Kategorien Poren-, Kluft-, Kluft-/Poren- und Kluft-/Karst-Grundwasserleiter Hydrogeologischen Übersichtskarte von Deutschland (HÜK250)
Reduziert	Hydraulische Leitfähigkeitsklasse des Aquifers	(-)	Nummerischer Zahlenwert. Die 12 hydraulischen Leitfähigkeitsklassen (in m/s) nach der Hydrogeologischen Übersichtskarte von Deutschland (HÜK250) wurden zu drei Klassen aggregiert. Stufe 1 (äußerst gering ($\leq 10^{-9}$), sehr gering ($>10^{-9} - 10^{-7}$), gering ($>10^{-7} - 10^{-5}$), gering bis äußerst gering ($\leq 10^{-5}$)) Stufe 2 (mittel ($>10^{-4} - 10^{-3}$), mittel bis mäßig ($>10^{-5} - 10^{-3}$), mäßig ($>10^{-5} - 10^{-4}$), mäßig bis gering ($>10^{-6} - 10^{-4}$), stark variabel) Stufe 3 (hoch ($>10^{-3} - 10^{-2}$), sehr hoch ($> 10^{-2}$))
Erweitert	Hydrogeologischer Großraum	(-)	Nummer des hydrogeologischen Großraumes ableitet nach der

Konfiguration	Feature	Einheit	Erläuterung
			Hydrogeologischen Raumgliederung Deutschlands (HYRAUM; BGR und SGD 2015)
Erweitert	Langjähriges Mittel der Lufttemperatur	(°C)	Langjähriges Mittel der Lufttemperatur (aus HYRAS Rasterdatensatz)
Erweitert	Mittlerer Sommerniederschlag	(mm)	Mittlerer Sommerniederschlag (aus HYRAS Rasterdatensatz)
Erweitert	Mittlerer Winterniederschlag	(mm)	Mittlerer Winterniederschlag (aus HYRAS Rasterdatensatz)
Erweitert	Verhältnis aus Niederschlag April-September/Okttober-März	(mm)	Verhältnis aus Niederschlag April-September/Okttober-März (aus HYRAS Rasterdatensatz)
Erweitert	Geländehöhe	(Meter über Meeresspiegel)	Entnommen aus (Amatulli et al. 2018).
Erweitert	Hangneigung	(°)	Entnommen aus Amatulli et al. (2018).
Erweitert	Gesteinsart	(-)	Gesteinsart differenziert nach sedimentären, magmatischen und metamorphen Gesteinen aus der Hydrogeologischen Übersichtskarte von Deutschland (HÜK250)
Erweitert	Geochemischer Typ	(-)	Die Informationen wurden der Hydrogeologischen Übersichtskarte von Deutschland (HÜK250) entnommen und beinhalten: silikatisch, silikatisch mit organischen Anteilen, silikatisch/karbonatisch, karbonatisch, organisch, sulfatisch, sulfatisch/halitisch, halitisch, unspezifisch, durch Auffüllung anthropogen verändert
Erweitert	DSD_1, DSD_2, LP_1, LP_2, SD_1, SD_2	(-)	Aus MOHP, abstrahierte Abstände zu Einzugsgebietsgrenzen und Gewässer für Strahler-Ordnung 1 und 2 (Nölscher et al. 2022)
Erweitert	Schluffgehalt im Boden	(%)	Tiefengewichteter Anteil des Schluffgehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
Erweitert	Tongehalte im Boden	(%)	Tiefengewichteter Anteil des Tongehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
Erweitert	Ackerfläche im Puffer (kulturunspezifisch)	(%)	Aus Sentineldaten abgeleitete Ackerfläche im Puffer der Messstelle (Blickensdörfer et al. 2022)

Konfiguration	Feature	Einheit	Erläuterung
Erweitert	Sickerwasserrate	(mm/Jahr)	Langjährige Sickerwasserrate pro Jahr (entnommen aus BGR, 2023)

A.3 Exkurs zum Einfluss des Klimawandels auf die Grundwasservulnerabilität durch PSM

Der Klimawandel stellt die menschliche Zivilisation vor schwer zu lösende Aufgaben. Seit Mitte des 20. Jahrhunderts ist eine signifikante Erwärmung der bodennahen Luftschichten über Land- und Meeresflächen zu verzeichnen. Der Klimawandel manifestiert sich zunehmend durch eine Häufung extremer Hitzeereignisse, einen fortschreitenden Anstieg des Meeresspiegels sowie durch regionale Veränderungen in der Häufigkeit und Intensität extremer Niederschlagsereignisse. Der wissenschaftliche Konsens weist auf anthropogene Einflüsse als Hauptursache dieser Entwicklungen hin (DWD 2025). Berücksichtigt man ausschließlich meteorologische Einflussfaktoren lässt sich ein Trend von absinkenden Grundwasserspiegeln – besonders in Norddeutschland – unter dem Repräsentative Konzentrationspfade (RCP) Szenario RCP8.5 bis zum Jahre 2100 feststellen (Wunsch et al. 2022). Dies geht mit einer erhöhten Saisonalität und längeren Phasen von tiefen Grundwasserspiegeln einher. Verschiebungen der Schnittstelle zwischen vadose (ungesättigter) und phreatischer (gesättigter) Zone sind also zu erwarten. Weiterhin haben Veränderungen des Klimas direkt Einfluss auf die Absatzmengen von PSM durch größere Schwankungen und individuelleres Auftreten von Schadorganismen (FAO 2023). Indirekt ist aber auch der Verbleib von PSM im Boden beeinflusst durch Veränderungen des Bodenwasserhaushaltes mit Einfluss auf die Translokation. PSM unterliegen weiterhin Transformationsprozessen, welche vorwiegend durch mikrobielle Aktivität (biotisch) gesteuert werden. Die Bodentemperatur nimmt daher indirekt Einfluss auf das Auftreten von Metaboliten und die Umsatzraten der Wirkstoffe, die potenziell ebenfalls in das Grundwasser verlagert werden können (Pérez-Lucas et al. 2024). Insbesondere die Funktionsleistung der Böden für den Abbau von PSM ist ein in der Zukunft schwer zu quantifizierender Faktor im Hinblick auf die Vulnerabilität des Grundwassers durch PSM.

Potenzielle Einflussfaktoren, die bei der zukünftigen Betrachtung und Bewertung von PSM im Grundwasser eine Rolle spielen, können wie folgt zusammengefasst werden (Bloomfield et al. 2006) (die folgende Liste bezieht sich auf Fallbeispiele und Klimaszenarien die für Großbritannien erstellt wurden):

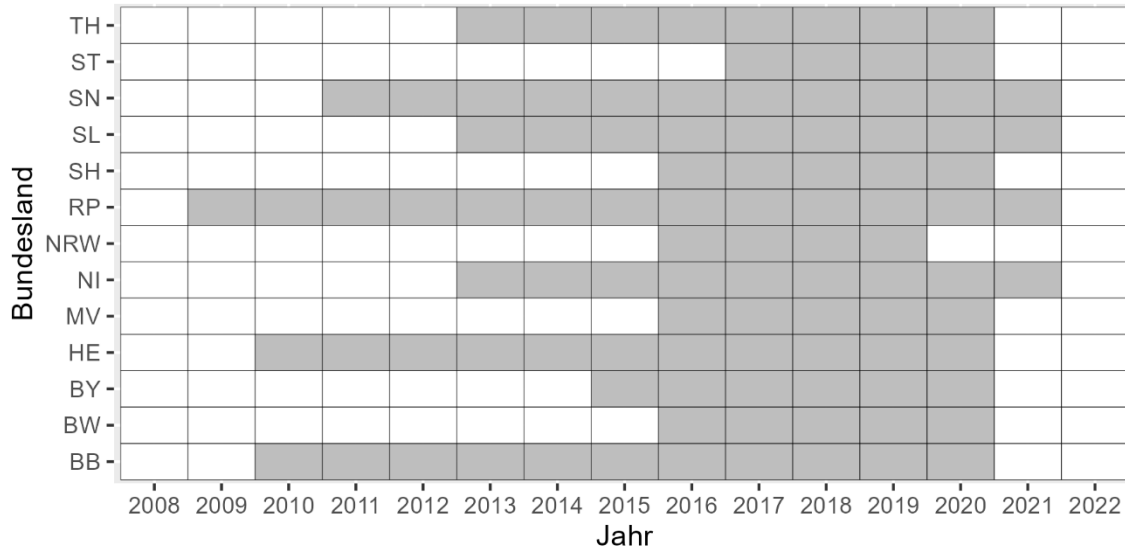
- ▶ *Intensivierung der Nahrungsmittelproduktion und Modifikation der landwirtschaftlichen Nutzfläche*
- ▶ *Veränderte Vegetationsperioden und Bedingungen die einen Anbau neuer Kulturfrüchte bedingen*
- ▶ *Erhöhte Winterniederschläge können Krankheiten bei Nutzpflanzen verstärken und dadurch PSM-Einsatz erhöhen*
- ▶ *Die organische Bodensubstanz (C_{org}) als wichtigster Sorbent für PSM unterliegt bei erhöhten Bodentemperaturen einen stärkeren Abbau. Zukünftige C_{org} Gehalte haben also indirekt Einfluss auf die Verlagerung von PSM*
- ▶ *Starkniederschlagsereignisse modifizieren den horizontalen Transport von PSM durch Erosion an der Bodenoberfläche und Austrag in Oberflächengewässer sowie den vertikalen Transport in der vadosen Zone*
- ▶ *Makroporenbildung (Ausmaß und Frequenz) in Böden nimmt bei steigenden Temperaturen und Modifikationen des Bodenwasserhaushaltes zu. Dies beeinflusst auf einer kleinen Maßstabsebene die Menge und Geschwindigkeit der Verlagerung von PSM*
- ▶ *Unkraut wird sich und veränderten Klimabedingungen vermutlich stärker durchsetzen, was mit einer Intensivierung von PSM-Anwendungen einhergehen kann*

- *Mildere Wintertemperaturen erhöhen den Druck durch Schädlingsbefall (Dauer und Häufigkeit des Auftretens) und auch das Auftreten in bisher nicht vulnerablen Regionen ist möglich*

Die hohe Vorhersagegüte für die Wirkstoff- und Metaboliten-Modelle, Konzentrationsklassen im Grundwasser vorherzusagen, haben sich vorwiegend aus dem Feature Landnutzung ergeben, i.e., der räumlich-zeitlich aggregierte Flächenanteil einer Kulturfrucht um die Grundwassermessstelle. Es ist daher auch anzunehmen, dass im Zuge des Klimawandels veränderte Vegetationsperioden und Bedingungen die zum Anbau von neuen Kulturfrüchten führen einen besonders starken Einflussfaktor für die Grundwasservulnerabilität durch PSM darstellen. Die Effekte, die sich durch einen Landnutzungswechsel auf die Absatzmenge von PSM ergeben, sind jedoch alles andere als einfach herauszuarbeiten. So führt z.B. das Brachfallen von landwirtschaftlicher Anbaufläche um 10% nicht zu einem unmittelbaren Rückgang der Absatzmenge von PSM, sondern sogar zu einem 0,6% höheren PSM-Absatz auf den benachbarten Flächen durch einen intensiveren Schädlingsdruck (Larsen et al. 2023). Innovationen (technisch, biologisch) in der Züchtung von widerstandsfähigen Kulturen, die zu einem Rückgang der Absatzmenge von PSM führen, sind ebenfalls schwer zu prognostizieren. Global gesehen geht eine Ertragssteigerung von 1% pro Hektar mit einem erhöhten PSM-Absatz von 1,6% pro Hektar einher, wobei dieser Effekt geringer ausfällt, je stärker wirtschaftlich ein Land aufgestellt ist (Schreinemachers et al. 2012).

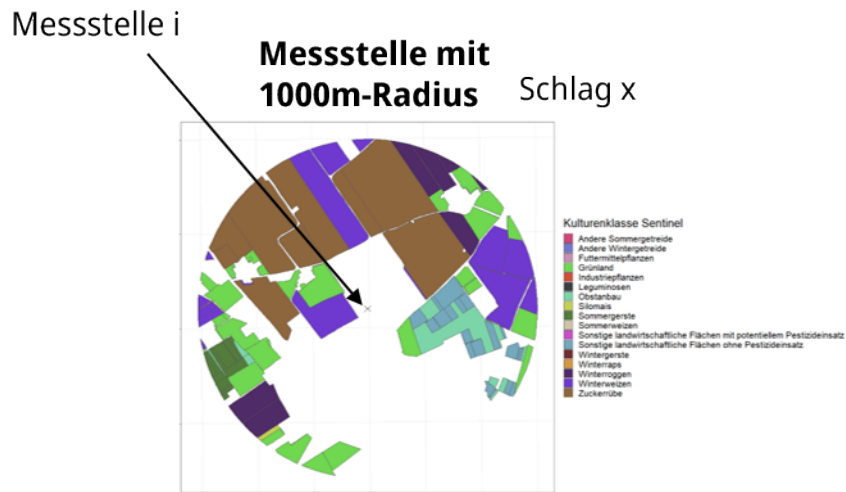
A.4 Zusatzinformationen zu InVeKoS-Daten.

Abbildung 127: Übersicht der im Projekt verfügbaren InVeKoS-Daten pro Jahr und Bundesland (grau = Daten verfügbar).



Quelle: eigene Darstellung BGR.

Abbildung 128: Beispiel von verarbeiteten InVeKoS-Daten, die den im Projekt verwendeten Kulturklassen zugeordnet und in einem 1-km-Radius um die Grundwassermessstelle ausgeschnitten wurden.



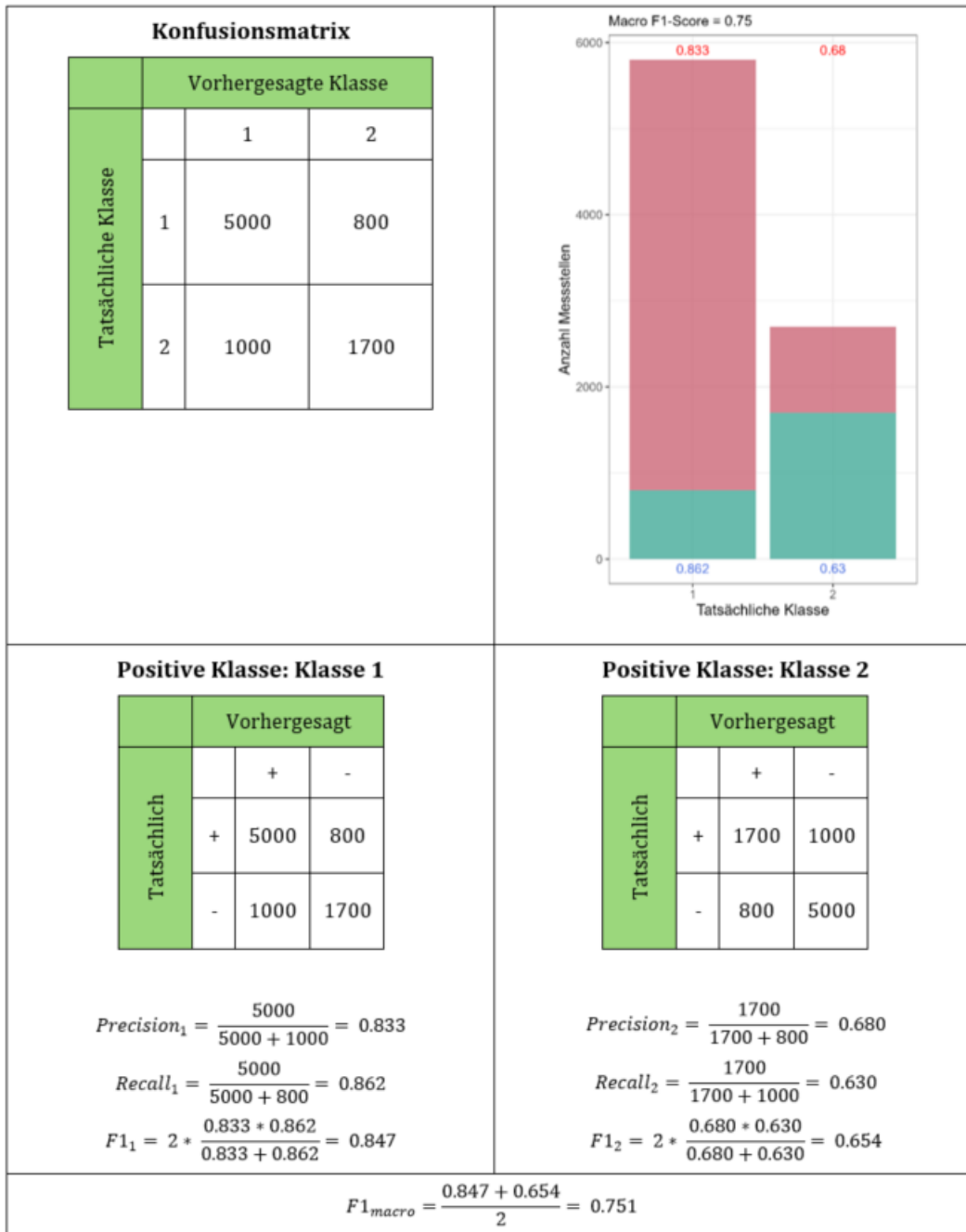
Kultur y auf Schlag x im Jahr j

Quelle: eigene Darstellung BGR.

A.5 Rechenbeispiele zur Bewertung der Modellgüte.

Abbildung 129: Rechenbeispiel für die klassenspezifische Ermittlung von Precision, Recall und F1-Score sowie die Berechnung des Macro-F1-Scores für eine binäre Klassifikation (2 Klassen). Für die Berechnungen werden vorhergesagte und tatsächliche Klasse zunächst in *One-vs-rest*-Konfusionsmatrizen umgewandelt. Erklärungen und Formeln zu den Berechnungen finden sich in 2.2.3.3.

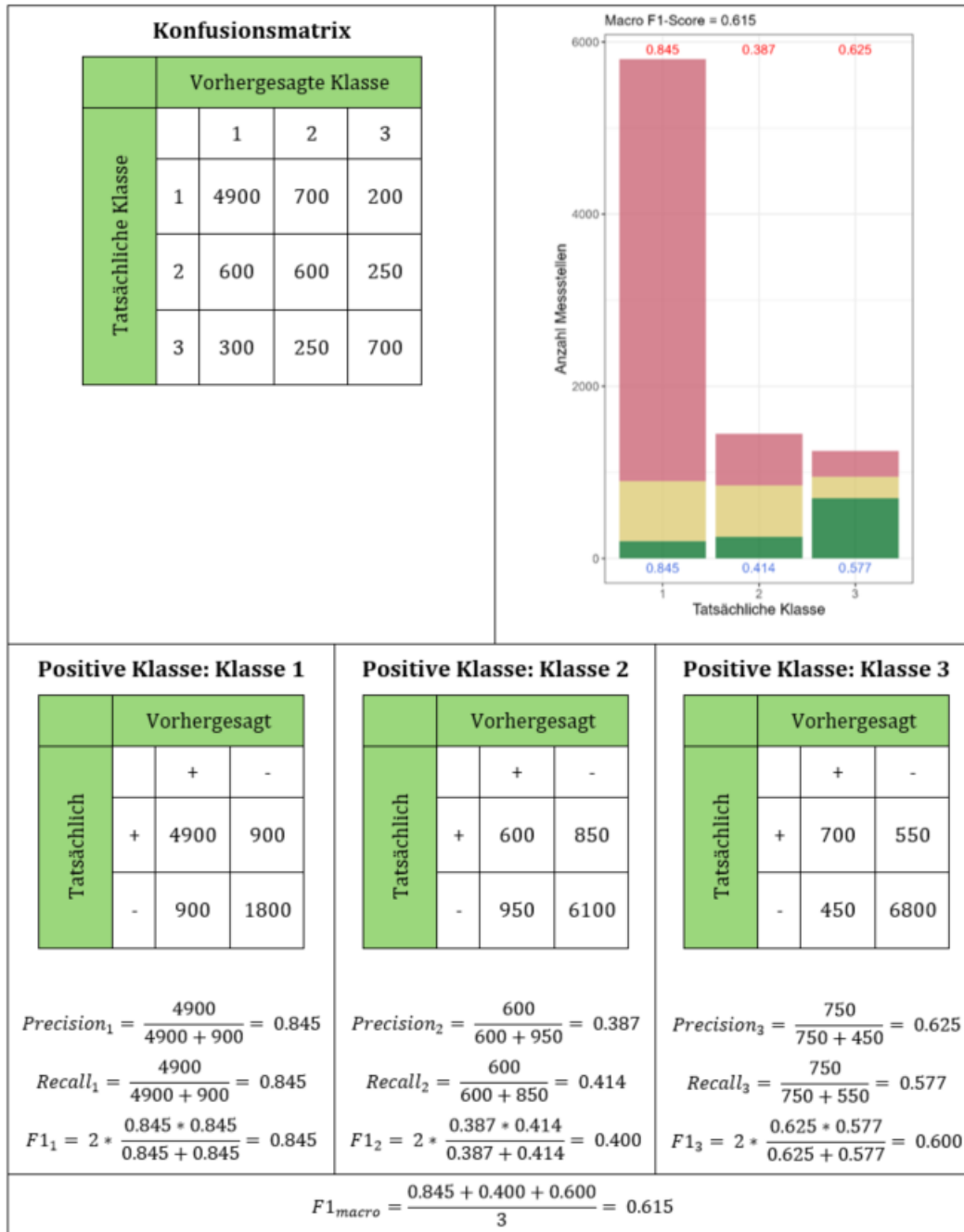
Binäre Klassifikation (2 Klassen)



Quelle: eigene Darstellung BGR.

Abbildung 130: Rechenbeispiel für die klassenspezifische Ermittlung von Precision, Recall und F1-Score sowie die Berechnung des Macro-Scores für eine Mehrklassen-Klassifikation (3 Klassen). Für die Berechnungen werden vorhergesagte und tatsächliche Klasse zunächst in *One-vs-rest*-Konfusionsmatrizen umgewandelt. Erklärungen und Formeln zu den Berechnungen finden sich in 2.2.3.3.

Mehrklassen-Klassifikation (3 Klassen)



Quelle: eigene Darstellung BGR.

A.6 Exkurs zu Pilotgebieten

Im ursprünglichen Projektplan wurde die Modellentwicklung anhand von Pilotgebieten skizziert. Letztendlich konnte die Entwicklung von robusten Modellen mit ausreichender Messstellenanzahl nur bundesweit gelingen. Weiterhin konnten im Datensatz keine Regionen bzw. Zusammenstellungen an Messstellen gefunden werden, an denen die Datenlage hinsichtlich aller Features vollständig ist. Die Modellentwicklung anhand von bundesweiten Daten musste zwar auf weniger hoch aufgelöste, dafür bundesweite Flächendaten zurückgreifen, konnte somit allerdings eine große Spannweite an Gegebenheiten abdecken, aus denen die Modelle lernen konnten. Ein Fokus auf zu kleinteilige Gebiete führt dazu, dass der Einfluss von bestimmten Standortfaktoren letztendlich nicht analysiert werden kann, da diese in diesem Teilbereich des Datensatzes nicht oder nicht ausreichend variieren. Es wurden Untersuchungen angestellt, inwiefern die Heterogenität der Datenverfügbarkeit in den Bundesländern insbesondere hinsichtlich der Höhe der Bestimmungsgrenzen eine Rolle spielt. Hierzu wurden Bundesländer als Pilotgebiete ausgewählt, die bei einer Auswahl an Stoffen eine deutlich geringere maximale Bestimmungsgrenze aufweisen als bundesweit der Fall ist. Somit sollte getestet werden, inwiefern Modelle in Pilotgebieten durch ggf. bessere Analytik eine bessere Vorhersagegüte erzielen könnten.

Es zeichnete sich ab, dass der Versuch, bundesweit trainierte Modelle mit Modellen, die anhand einer regionalen Messstellenauswahl mit unterschiedlichen Bestimmungsgrenzen trainiert wurden, miteinander vergleichen zu wollen nicht sinnvoll ist. Unterschiede lassen sich nicht zwangsläufig auf den Einfluss niedriger Bestimmungsgrenzen beziehen. Das liegt bereits an der unterschiedlichen Menge der Daten bzw. der Anzahl der Messstellen, aber auch am Bias der unterschiedlichen Anbaugebiete mit ihren unterschiedlichen Eigenschaften. Deutlich sinnvoller ist der Vergleich von zwei Modellen, die beide auf denselben Daten derselben Region trainiert worden sind, sich aber in der Aufteilung der Konzentrationsklassen unterscheiden. Regional sollen Metaboliten-Modelle mit der bundesweiten Konzentrationsklassengrenze $0,05 \mu\text{g L}^{-1}$ und länderspezifische Modelle mit deutlich niedrigerer maximaler Bestimmungsgrenze (z.B. $0,01 \mu\text{g L}^{-1}$ in Sachsen) mit einander verglichen werden.

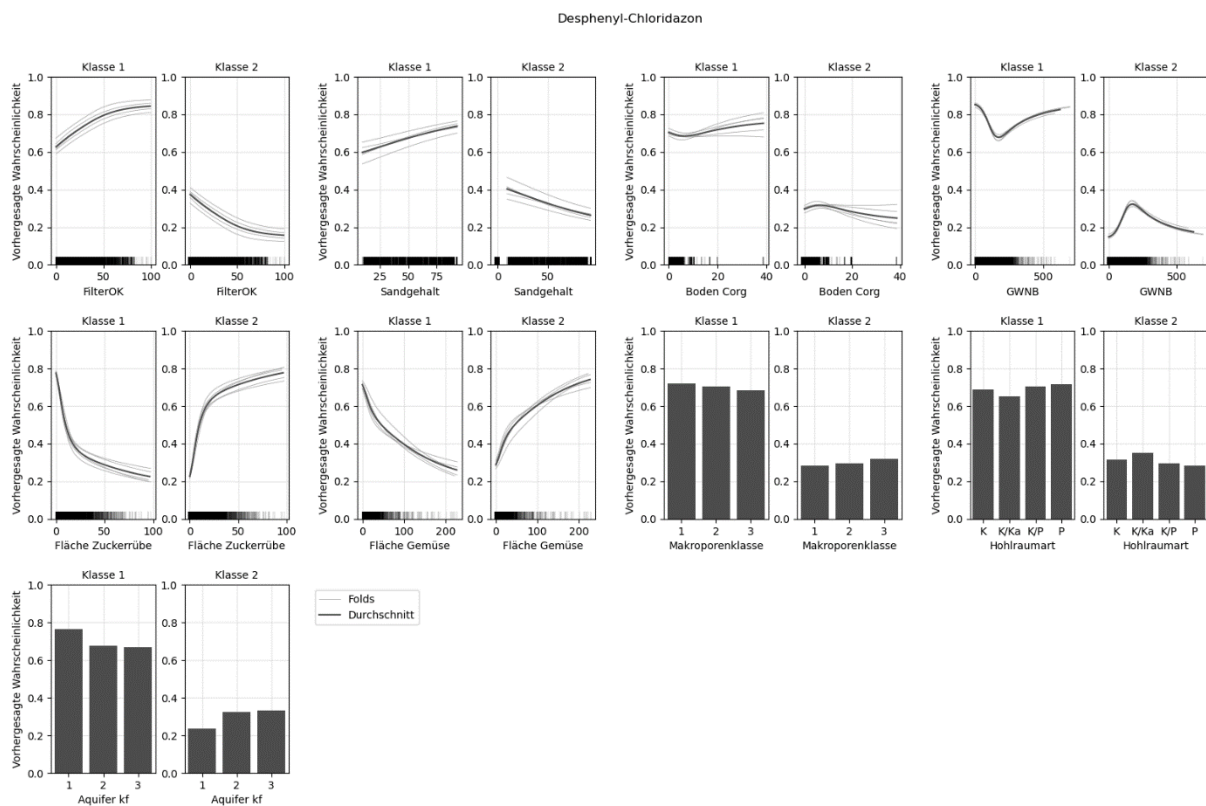
Hierzu wurden Modelle für die in den Projektergebnissen gezeigten Metaboliten in den Bundesländern mit unterschiedlichen Klassengrenzen getestet. Für die meisten Stoffe in den Bundesländern, die getestet wurden, gab es keine eindeutigen Verbesserungen. Der Macro-F1-Score unterschied sich nur geringfügig zwischen einem Modell für Metazachlor ESA in Sachsen mit der Klassengrenze $0,05 \mu\text{g L}^{-1}$ und einem Modell mit der Klassengrenze $0,01 \mu\text{g L}^{-1}$, bei einem Vergleich in Thüringen wurde eine leichte Verschlechterung festgestellt. Bei einem Versuch für Desphenyl-Chloridazon in Sachsen waren leichte Verbesserungen sichtbar. Eine weitere Unsicherheit bei der Interpretation der Performance sind sicherlich auch die wechselnden Stärken der Ungleichverteilung mit der Klassengrenze. Es ist z.B. schwierig zu sagen, ob eine bessere Performance nur durch die besser Gleichverteilung der Klassen zustande kommt (da f1-Score das harmonische Mittel ist). Bei den untersuchten Beispielen kommt es nicht zu einer pauschalen Verbesserung der Performance, sodass aus den Pilotgebieten keine eindeutigen Schlüsse zur Hypothese von verbesserten Modellen durch bessere Analytik gezogen werden können. Eine bessere Gleichverteilung zwischen den Klassen führt also nicht pauschal zur Verbesserung der Performance. Vielmehr ist je nach Fragestellung die Wahl der Klassengrenzen wichtig. Da die Modelle in Pilotgebieten nur für mobile Wirkstoffe mit langer Anwendungshistorie verglichen werden konnten, sind für diese Stoffe geringere Bestimmungsgrenzen nicht unbedingt notwendig. Wie aus der Diskussion der Modellergebnisse im Bericht ersichtlich wird, wird davon ausgegangen, dass diese hoch mobilen Metaboliten einen

Großteil der Messstellen erreicht haben und dass andere Klassengrenzen bei der Interpretation dieser Modelle nicht unbedingt hilfreich sind.

Dennoch würde eine bessere und einheitliche Analytik Unsicherheiten ausräumen und zu mehr Spielraum bei der Wahl der Klassengrenzen führen, auch um diese These überhaupt prüfen zu können. Dies gilt insbesondere für relevante Metaboliten und Wirkstoffe, da die für dieses Modell verwendeten nichtrelevanten Metaboliten häufig schon sehr lange gebildet werden und durch ihre hohe Mobilität und Persistenz in hohen Konzentrationen vorkommen. Zusammen mit Daten zu tatsächlich aufgetragenen Wirkstoffmengen könnte durchaus mehr über die Vulnerabilität von Standorten gelernt werden.

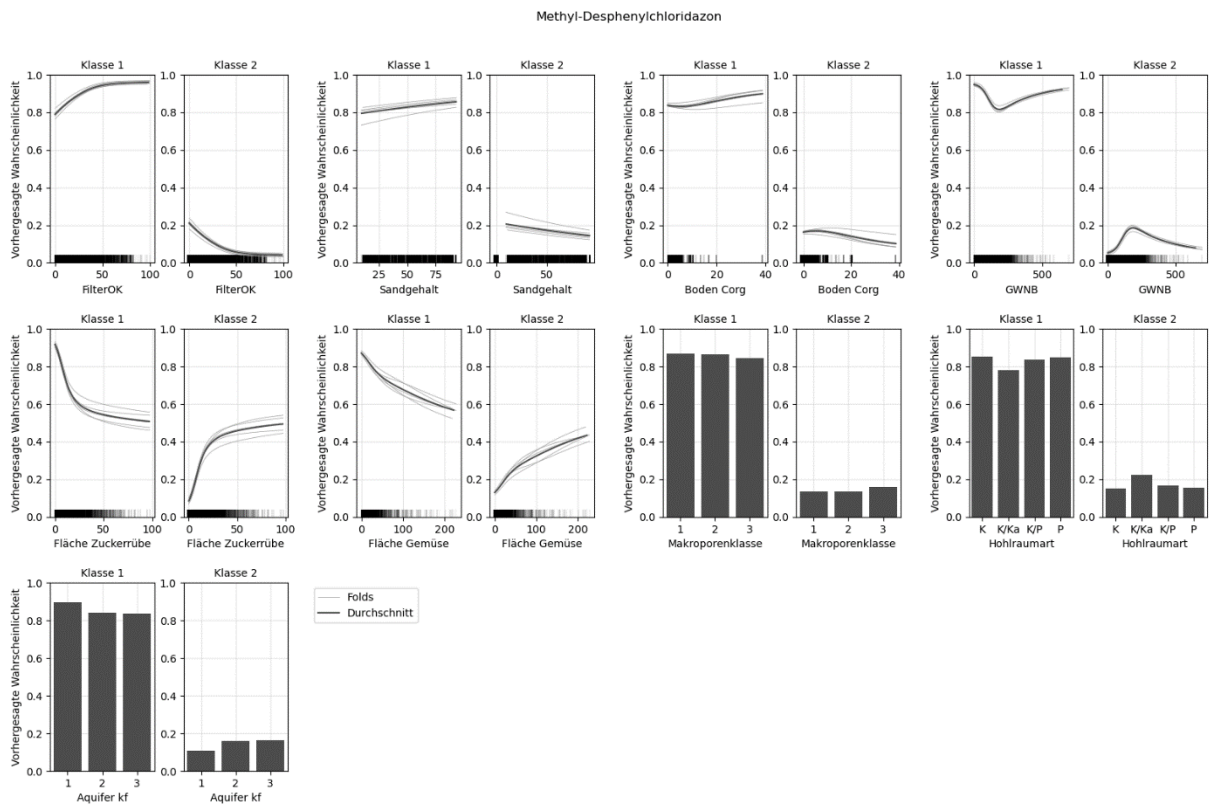
A.7 Feature-Effekte der reduzierten Feature-Auswahl in den Neuronalen Netze

Abbildung 131: Partial-Dependence-Plots für Desphenyl-Chloridazon.



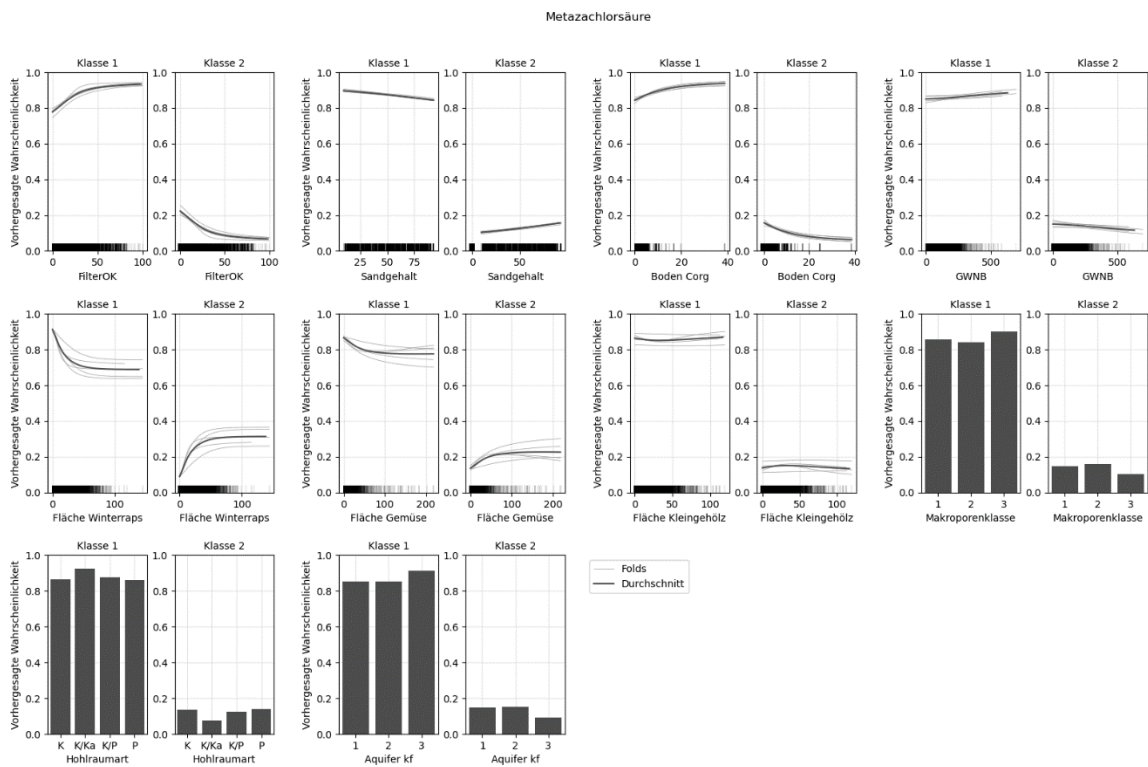
Quelle: eigene Darstellung BGR.

Abbildung 132: Partial-Dependence-Plots für Methyldesphenyl-Chloridazon.



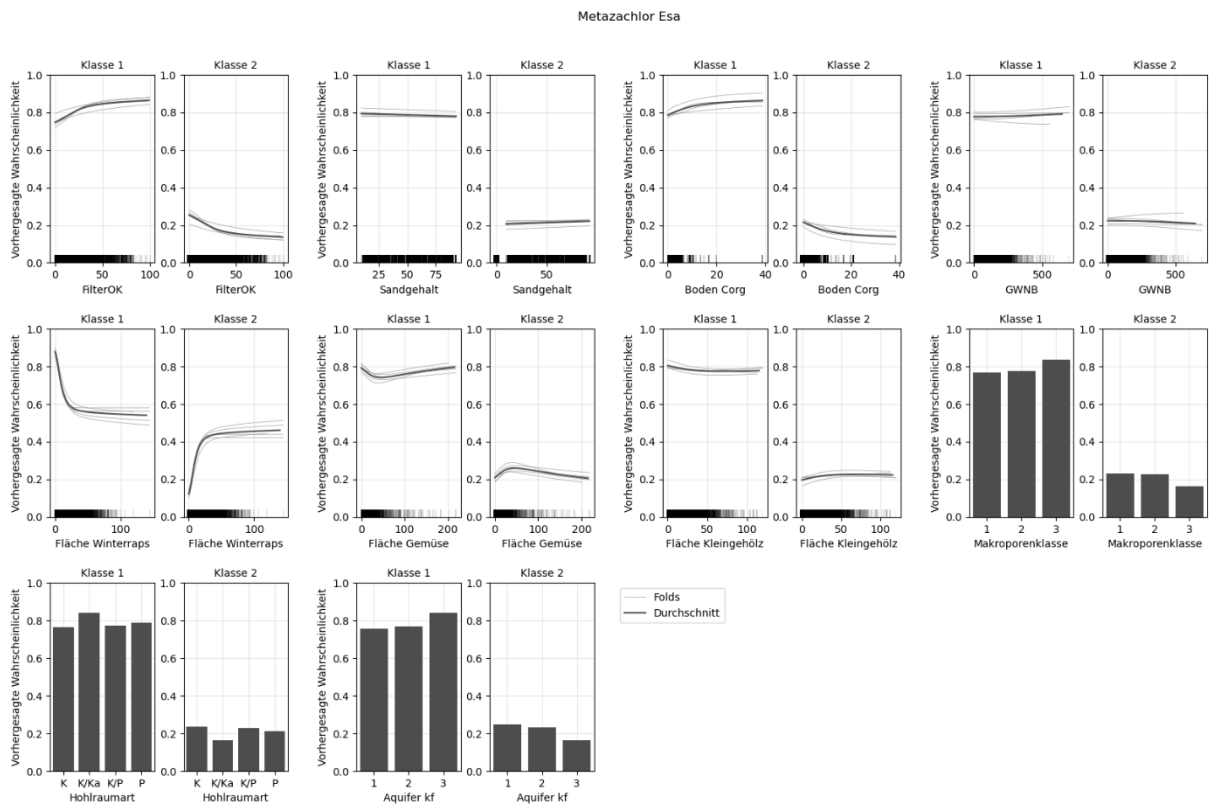
Quelle: eigene Darstellung BGR.

Abbildung 133: Partial-Dependence-Plots für die Metazachlorsäure.



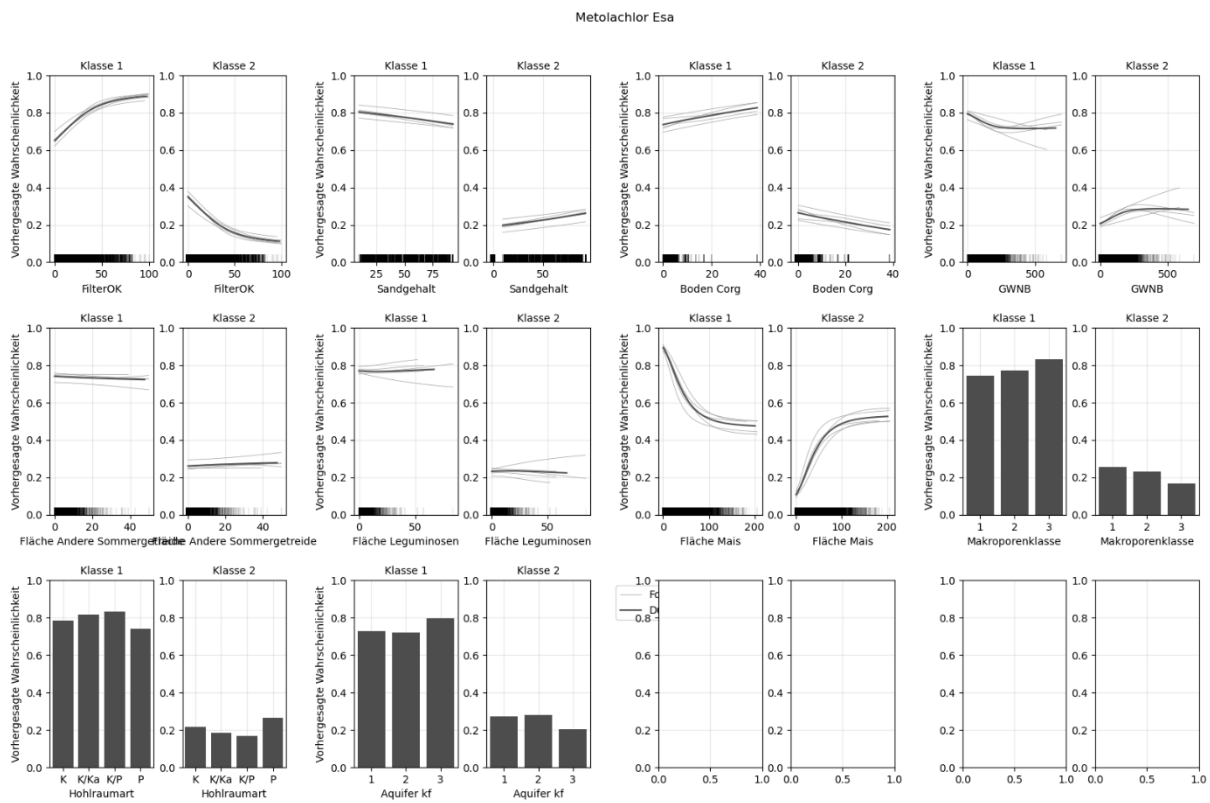
Quelle: eigene Darstellung BGR.

Abbildung 134: Partial-Dependence-Plots für Metazachlor ESA.



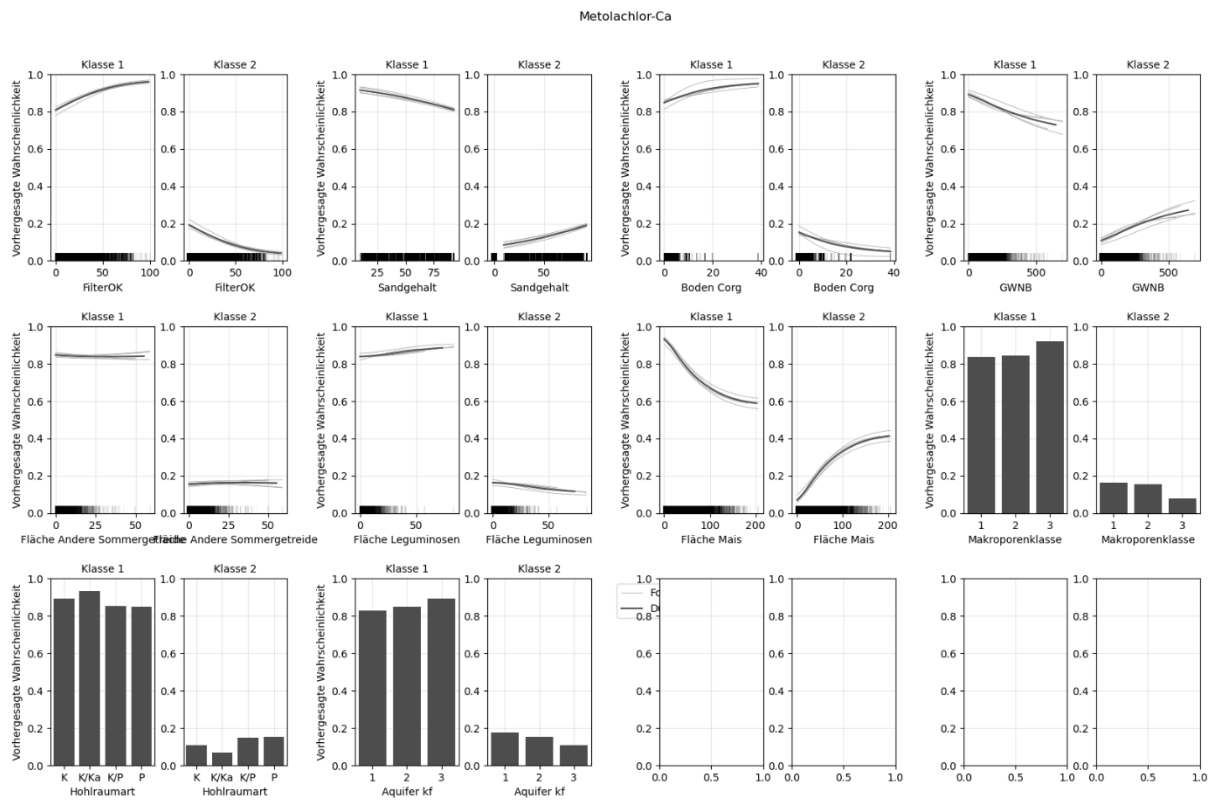
Quelle: eigene Darstellung BGR.

Abbildung 135: Partial-Dependence-Plots für Metolachlor ESA.



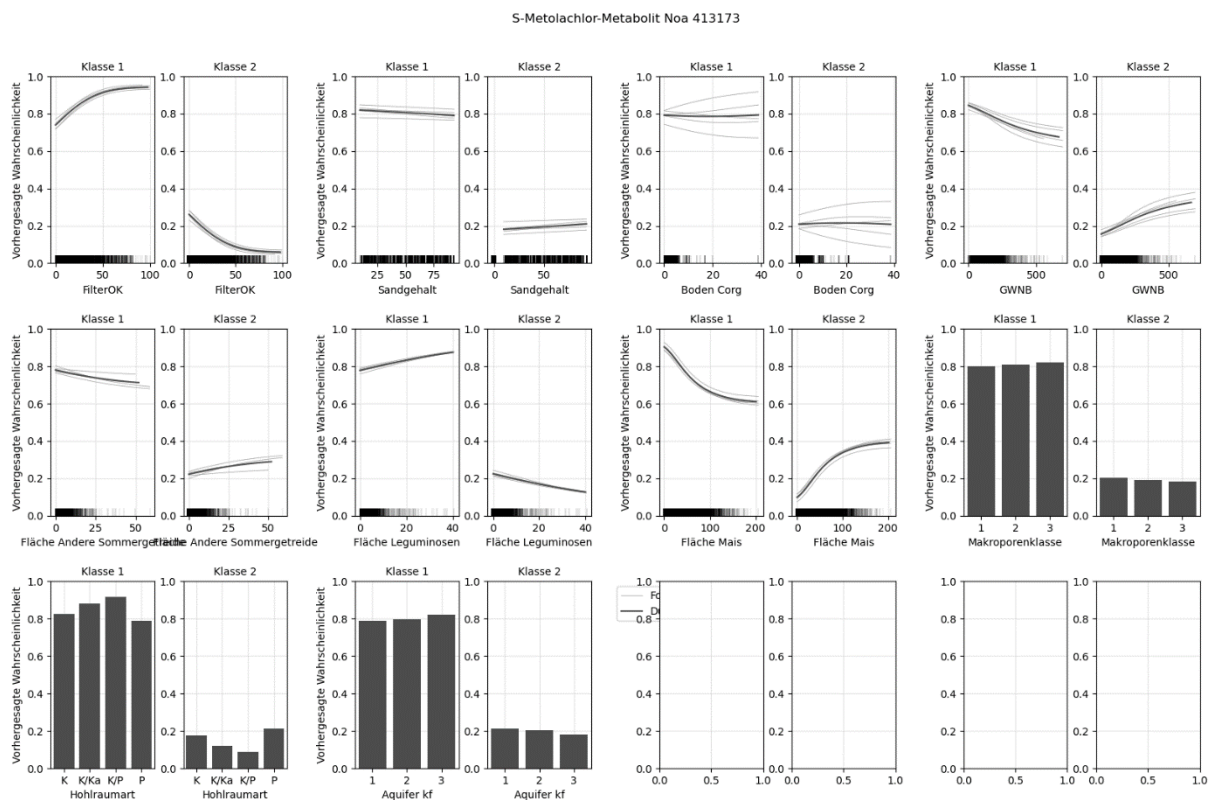
Quelle: eigene Darstellung BGR.

Abbildung 136: Partial-Dependence-Plots für Metolachlor OXA.



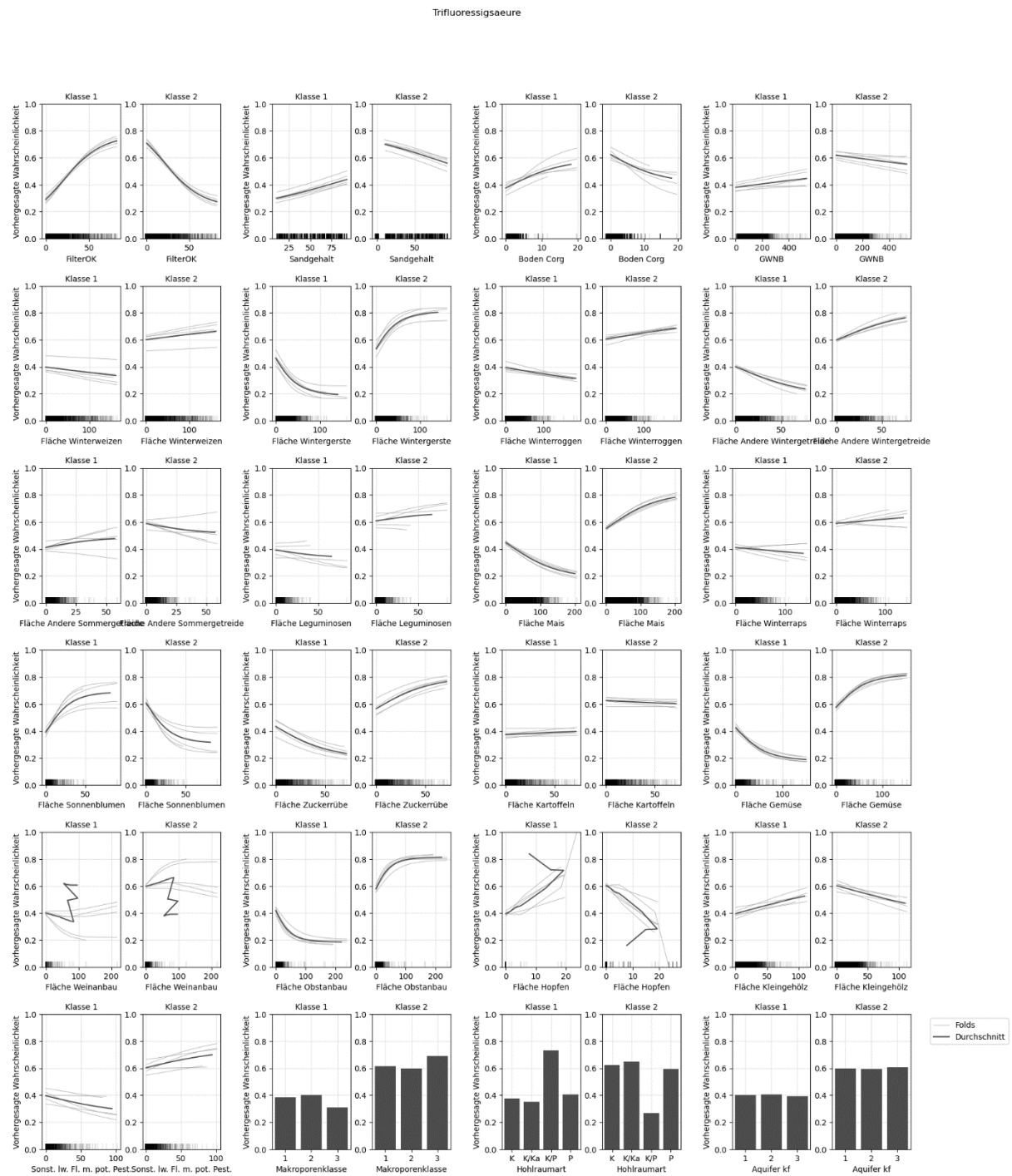
Quelle: eigene Darstellung BGR.

Abbildung 137: Partial-Dependence-Plots für Metolachlor Metabolit NOA 413173.



Quelle: eigene Darstellung BGR.

Abbildung 138: Partial-Dependence-Plots für Trifluoressigsäure.



Quelle: eigene Darstellung BGR.

A.8 Zusammenstellung von Kriterien die für eine Vielzahl von Verfahren für die Vulnerabilitätsbewertung von Grundwasser herangezogen werden.

Tabelle 9: Methodenübersicht zu Verfahren der Vulnerabilitätsbewertung die in der Grundwasserbewertung angewendet wurden.

Quelle	C _{org}	BA	FK	Tex	m _{Bo}	K _{fBo}	Lit _{vad}	m _{vad}	K _{fvad}	T _{res}	GW _{Ab}	GW _n	K _{fphr}	Art _{phr}	m _{phr}	OFW/GW	Lit _{phr}	LN	To	Weitere
(Albinet et al. 1970)							X	X			X			X		X	X			
(Haertle et al. 1982)							X	X			X									
(Rao et al. 1985)	X	X	X								X	X								Pestizidmengen und Halbwertszeiten, Bodendichte
TGL 34334 (1986)								X			X			X			X			
(Carter et al. 1987)	X			X		X	X	X												
(Aller et al. 1987)		X					X				X	X	X	X					X	
(Foster 1987)							X	X			X			X			X			
(Voigt 1987)							X	X			X				X		X			Verwerfungen, sofern verfügbar
(McRae 1989)				X			X				X								X	
(Fobe et al. 1990)							X	X			X						X			
(Roeper 1990)						X	X	X												
(Carpenter 1992)	X			X	X	X	X				X			X			X		X	pH-Werte Boden
(Van Stempvoort et al. 1993)					X	X		X	X											
(Ray et al. 1993)									X			X				X			X	
(Sokol et al. 1993)	X	X	X									X							X	Vegetationsperioden
(Seelig 1994)	X					X			X		X									Kfoc und Halbwertszeit des Pestizids

Quelle	C _{org}	BA	FK	Tex	m _{Bo}	K _{fBo}	Lit _{vad}	m _{vad}	K _{fvad}	T _{res}	GW _{Ab}	GW _n	K _{fphr}	Art _{phr}	m _{phr}	OFW/GW	Lit _{phr}	LN	To	Weitere
(Zelykman et al. 1994)	X	X		X	X		X	X												pH-Werte Boden
(Bencini et al. 1995)							X	X	X											
(Höltling et al. 1995)			X				X	X			X	X		X						
(Civita et al. 1995)		X			X		X	X			X	X		X	X			X	X	
(Meinardi et al. 1995)	X	X						X		X		X					X	X		
(Scharp et al. 1997)											X	X	X		X					Grundwasserqualität und Nutzung
(Maxe et al. 1998)										X	X									
(Daly et al. 1998)		X		X	X	X	X	X	X	X	X	X	X	X		X	X			Art des Eintrags (diffus, Punktquelle)
(DELG 1999)		X		X	X	X	X	X	X	X	X	X	X	X		X	X			
(Fitzsimons et al. 2003)		X		X	X	X	X	X	X	X	X	X	X	X		X	X			
(Mádl-Szonyi et al. 1998)		X			X		X	X			X	X	X			X		X		hydraulischer Gradient, tektonische Elemente (falls benötigt)
(Malik et al. 1998)					X						X		X		X		X	X		Geomorphologie und Fließgeschwindigkeiten
(Palmer et al. 1998)	X			X			X	X					X				X			
(Rine et al. 1998)						X	X	X												
(Doerfliger et al. 1999)					X	X		X	X			X				X		X	X	Geomorphologie für Epikarst Identifizierung (falls benötigt)
(Goldscheider et al. 2000)						X	X	X	X			X		X		X			X	
(Heinkele et al. 2002)		X	X				X	X		X	X	X	X				X			
(LAWA 2003b)							X	X				X								Druckverhältnisse
(Albert 2005)							X	X				X	X	X	X		X	X		

Quelle	C _{org}	BA	FK	Tex	m _{Bo}	K _{fBo}	Lit _{vad}	m _{vad}	K _{fvad}	T _{res}	GW _{Ab}	GW _n	K _{fphr}	Art _{phr}	m _{phr}	OFW/GW	Lit _{phr}	LN	To	Weitere	
(Chachadi et al. 2005)											X		X	X	X					Entfernung von Küste, Verhältnis Cl ⁻ /(HCO ₃ ⁻ + CO ₃ ²⁻)	
(Mardhel et al. 2006)								X											X	Geomorphologische Indikatoren	
(Vias et al. 2006)				X	X		X	X				X		X		X					
(Ravbar et al. 2007)		X		X	X		X	X		X		X	X							X	Geomorphologie
(Kavouri et al. 2011)				X	X		X	X	X								X			X	Epikarst Charakterisierung
(Wirsing et al. 2015)	X	X	X				X	X	X		X	X		X							effektive Wurzelzone, pH Boden

Die Quellen und die darin enthaltenen Methoden sind aufsteigend sortiert nach Datum der Veröffentlichung.

Abkürzungsverzeichnis der Tabellenüberschriften: Organischer Kohlenstoffgehalt des Boden = C_{org}; Bodenart = BA; Feldkapazität = FK; Textur = Tex; Bodenhorizontmächtigkeit = m_{Bo}; Hydraulische Leitfähigkeit des Bodens = K_{fBo}; Lithologie der vadosen Zone = Lit_{vad}; Mächtigkeit der vadosen Zone = m_{vad}; Hydraulische Leitfähigkeit der vadosen Zone = K_{fvad}; Verweilzeit in der vadosen Zone = T_{res}; Grundwasserflurabstand = GW_{Ab}; Grundwasserneubildung = GW_n; Hydraulische Leitfähigkeit des Aquifers = K_{fphr}; Aquiferart = Art_{phr}; Aquifermächtigkeit = m_{phr}; Interaktion zwischen Oberflächengewässer und Grundwasser = OFW/GW Interaktion; Lithologie = Lit; Landnutzung = LN; Topographische Informationen = To

Bodenparameter
Ungesättigte Zone
Aquifer Parameter
Weitere Parameter
Zusätzlich benötigte Daten