

# Dokumentation des Datensatzes zum FARM-Projekt (FKZ 3720234010)

Stand: 25.09.2025

Der im folgenden beschriebene Datensatz wurde im Forschungs- und Entwicklungsvorhaben mit dem Förderkennzeichen 3720234010 des Umweltbundesamtes durch die Bundesanstalt für Geowissenschaften und Rohstoffe bearbeitet. Die Vorstellung der Ergebnisse und die Diskussion dieser sind ausführlich im Endbericht mit dem Titel „Grundwasservulnerabilitätsbewertung im ZulassungsverFAhRen von PflanzenschutzMitteln (FARM)“ veröffentlicht und auf der Seite des Umweltbundesamtes zu finden (Cooke et al. 2025). Der Endbericht ist voraussichtlich ab dem vierten Quartal 2025 auf der Publikationsbibliothek des Umweltbundesamtes zu finden.

Der Datensatz beinhaltet vier .csv-Dateien:

- ▶ Stammdaten der Grundwassermessstellen = *stammdaten\_geo.csv*
- ▶ PSM Wirkstoff- und Metabolitdaten = *psm\_messwerte.csv*
- ▶ Grundwasserchemische Daten = *gw\_chemie\_messwerte.csv*
- ▶ Aggregierte Informationen zu Grundwasserflurabständen = *gw\_flab\_stat.csv*

Diese Daten wurden zum Trainieren der Modelle des maschinellen Lernens verwendet. Die Informationen der Stammdaten der Grundwassermessstellen wurden zusammen mit den hydrologischen Daten zur Klassifikation von potenziell kompatiblen Messstellen genutzt, die sich zum Monitoring im Zulassungsverfahren von Pflanzenschutzmitteln heranziehen lassen.

## Stammdaten und verknüpfte Geodaten (features) für das maschinelle Lernen

Bezeichnung	Erläuterung
messstellen_id	Nummerierung der Messstellen, entspricht der ID aus dem jeweiligen Bundesland, in manchen Fällen wurden die originalen Messstellen IDs anonymisiert oder modifiziert (siehe Disclaimer in diesem Dokument)
bundesland	Das Bundesland aus dem die Daten bereitgestellt wurden definiert über das Bundeslandkürzel ("BB", "BW", "BY", "HE", "MV", "NI", "NRW", "RP", "SH", "SL", "SN", "ST", "TH")
messnetz	Informationen zum Messnetz (bspw., Landesmessnetz, Wasserversorger, Daten Dritter, keine Information)
filter_ok_unter_gok	Filteroberkante in Meter unter Geländeoberkante
filter_uk_unter_gok	Filterunterkante in Meter unter Geländeoberkante
gok_m_nn_	Geländeoberkante
messpunkt-höhe_m_nn_	Messpunkthöhe
end-teufe_m_u_an-satzpunkt_	Endteufe
erfassungsdatum_YYYY_	Jahr der Erfassung der Messstelle im jeweiligen Länder-Datensatz
messpunkthöhe	Messpunkthöhe
messstellenart	Informationen zur Art der Messstelle (bspw., Brunnen, Stollen, Quelle, Schüttung, ...)
messstellen_id2	Zweite Messstellen-ID, falls vorhanden
filtermaterial_verein-facht	Material des Filters (bspw., PVC, HDPE, Stahl, ...)
x	Rechtswert
y	Hochwert
epsg	Schlüsselnummern (EPSG-Code) für Koordinatenreferenzsysteme und geodätische Datensätze, in diesem Fall EPSG 25832 (Name: ETRS89 / UTM Zone 32N)

area_sentinel_id_1	Flächenanteil der Kulturklasse <i>Winterweizen</i> im Puffer der Messstelle (m <sup>2</sup> ) (Sentinel Daten wurden aus (Blickensdörfer et al. 2022) entnommen)
area_sentinel_id_2	Flächenanteil der Kulturklasse <i>Wintergerste</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_3	Flächenanteil der Kulturklasse <i>Winterroggen</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_4	Flächenanteil der Kulturklasse <i>Andere Wintergetreide</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_5	Flächenanteil der Kulturklasse <i>Sommergerste</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_6	Flächenanteil der Kulturklasse <i>Sommerhafer</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_7	Flächenanteil der Kulturklasse <i>Andere Sommergetreide</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_8	Flächenanteil der Kulturklasse <i>Grünland</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_9	Flächenanteil der Kulturklasse <i>Leguminosen</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_10	Flächenanteil der Kulturklasse <i>Mais</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_12	Flächenanteil der Kulturklasse <i>Winterraps</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_13	Flächenanteil der Kulturklasse <i>Sonnenblumen</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_14	Flächenanteil der Kulturklasse <i>Zuckerrübe</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_15	Flächenanteil der Kulturklasse <i>Kartoffeln</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_16	Flächenanteil der Kulturklasse <i>Gemüse</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_17	Flächenanteil der Kulturklasse <i>Weinbau</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_18	Flächenanteil der Kulturklasse <i>Obstanbau</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_19	Flächenanteil der Kulturklasse <i>Hopfen</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_20	Flächenanteil der Kulturklasse <i>Kleingehölz</i> im Puffer der Messstelle (m <sup>2</sup> )
area_sentinel_id_21	Flächenanteil der Kulturklasse <i>Sonstige landwirtschaftliche Flächen mit potentiellm Pestizideinsatz</i> im Puffer der Messstelle (m <sup>2</sup> )
sand_depth_weighted_mean	Tiefengewichteter Anteil des Sandgehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante (%)
makroporen_rounded	Differenzierung der Klassen: no macropore flow (1), moderate macropore flow (2), high macropore flow (3); (Klein et al. 2025)
gwn_mean	Mittlere langjährige Rate der Grundwasserneubildung (BGR 2019) (mm/a)
HA	Hohlraumart des Aquifers. Differenzierung der Kategorien Poren-, Kluft-, Kluft-/Poren- und Kluft-/Karst-Grundwasserleiter Hydrogeologischen Übersichtskarte von Deutschland (HÜK250)
kf_be	Hydraulische Leitfähigkeitsklasse des Aquifers. Numerischer Zahlenwert. Die 12 hydraulischen Leitfähigkeitsklassen (in m/s) nach der Hydrogeologischen Übersichtskarte von Deutschland (HÜK250) wurden zu drei Klassen aggregiert. <b>Stufe 1</b> (äußerst gering ( $\leq 10^{-9}$ ), sehr gering ( $>10^{-9} - 10^{-7}$ ), gering ( $>10^{-7} - 10^{-5}$ ), gering bis äußerst gering ( $\leq 10^{-5}$ )) <b>Stufe 2</b> (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) <b>Stufe 3</b> (hoch ( $>10^{-3} - 10^{-2}$ ), sehr hoch ( $> 10^{-2}$ ))
GR_NR	Nummer des hydrogeologischen Großraumes abgeleitet nach der Hydrogeologischen Raumgliederung Deutschlands (HYRAUM; BGR und SGD 2015)
airtemp_mean_mean_Decade_2010_2019	Langjähriges Mittel der Lufttemperatur (aus HYRAS Rasterdatensatz; °C)
mean_precip_summer	Mittlerer Sommerniederschlag (aus HYRAS Rasterdatensatz; mm)
mean_precip_winter	Mittlerer Winterniederschlag (aus HYRAS Rasterdatensatz; mm)
precip_ratio_as_om	Verhältnis aus Niederschlag April-September/Okttober-März (aus HYRAS Rasterdatensatz; mm)
elevation_elevation_1KMmd	Entnommen aus (Amatulli et al. 2018).
slope_slope_1KMmd	Entnommen aus Amatulli et al. (2018).
GA	Gesteinsart differenziert nach sedimentären, magmatischen und metamorphen Gesteinen aus der Hydrogeologischen Übersichtskarte von Deutschland (HÜK250)
GC	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

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)
schluff_depth_weighted_mean	Tiefengewichteter Anteil des Schluffgehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
ton_depth_weighted_mean	Tiefengewichteter Anteil des Tongehaltes abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante
percent_nutzung_Acker	Aus Sentineldaten abgeleitete Ackerfläche im Puffer der Messstelle (Blickensdörfer et al. 2022)
swr_mean	Langjährige Sickerwasserrate pro Jahr (entnommen aus BGR, 2023)
corg_gehalt_depth_weighted_mean	Tiefengewichteter Anteil an organischen Kohlenstoff abgeleitet aus der Bodenübersichtskarte 1:250.000 bis 1 m unter Bodenoberkante

## PSM Wirkstoff- und Metabolitdaten

Bezeichnung	Erläuterung
messstellen_id	Nummerierung der Messstellen, entspricht der ID aus dem jeweiligen Bundesland, in manchen Fällen wurden die originalen Messstellen IDs anonymisiert oder modifiziert (siehe Disclaimer in diesem Dokument)
bundesland	Das Bundesland aus dem die Daten bereitgestellt wurden definiert über das Bundeslandkürzel ("BB", "BW", "BY", "HE", "MV", "NI", "NRW", "RP", "SH", "SL", "SN", "ST", "TH")
lawa_name	Stoffnamen nach LAWA-Nomenklatur (falls fehlend ggf. eigens hinzugefügt)
lawa_nr	Stoffnummern nach LAWA-Nomenklatur (falls fehlend ggf. eigens hinzugefügt)
obs_mw	Messwerte und Werte unter der Bestimmungsgrenze bzw. Nachweisgrenze. Um was es sich handelt, ist den zwei nachfolgenden Spalten zu entnehmen.
censored_mw	TRUE bei zensiert (Wert unter Bestimmungs- bzw. Nachweisgrenze) FALSE bei nicht-zensiert (es handelt sich um einen Messwert)
class	MW steht für Messwert, BG für Bestimmungsgrenze, NG für Nachweisgrenze
einheit	Stoffkonzentration einheitlich in µg/L
medium	Informationen zur Herkunft der Grundwasserprobe
probenahmetiefe	Abstand in Meter unter Geländeoberkante
jahr	Jahr in dem die Messung stattfand
monat	Monat in dem die Messung stattfand
tag	Tag an dem die Messung stattfand
date	Zeitstempel zusammengesetzt aus den Spalten jahr-monat-tag
messnetz	Informationen zum Messnetz (bspw., Landesmessnetz, Wasserversorger, Daten Dritter, keine Information)

## Grundwasserchemische Daten

Die Bezeichnungen sind identisch zum Datenblatt mit den PSM Wirkstoff und Metabolitkonzentrationen aber bei den grundwasserchemischen Daten umfasst die Spalte lawa\_name folgende Informationen:

- ▶ aluminium
- ▶ ammonium
- ▶ arsen
- ▶ bor
- ▶ calcium
- ▶ chlorid
- ▶ organischer kohlenstoff, gelöst

- ▶ eisen
- ▶ elektrische leitfähigkeit bei 25°C
- ▶ fluorid
- ▶ gesamthärte
- ▶ hydrogencarbonat
- ▶ kalium
- ▶ kupfer
- ▶ magnesium
- ▶ mangan
- ▶ natrium
- ▶ nitrat-stickstoff
- ▶ nitrat
- ▶ ph-wert
- ▶ phosphor, gesamt
- ▶ orthophosphat-phosphor
- ▶ redox-spannung
- ▶ sauerstoff
- ▶ selen
- ▶ silizium
- ▶ sulfat
- ▶ wassertemperatur
- ▶ zink

## Aggregierte Informationen zu Grundwasserflurabständen

Bezeichnung	Erläuterung
messstellen_id	Nummerierung der Messstellen, entspricht der ID aus dem jeweiligen Bundesland, in manchen Fällen wurden die originalen Messstellen IDs anonymisiert oder modifiziert (siehe Disclaimer in diesem Dokument)
bundesland	Das Bundesland aus dem die Daten bereitgestellt wurden definiert über das Bundeslandkürzel ("BB", "BW", "BY", "HE", "MV", "NI", "NRW", "RP", "SH", "SL", "SN", "ST", "TH")
messnetz	Informationen zum Messnetz (bspw., Landesmessnetz, Wasserversorger, Daten Dritter, keine Information)
mean	Mittlerer Grundwasserflurabstand der Zeitreihe in Metern unter Geländeoberkante
min	Geringster Grundwasserflurabstand der Zeitreihe in Metern unter Geländeoberkante
max	Höchster Grundwasserflurabstand der Zeitreihe in Metern unter Geländeoberkante
sd	Standardabweichung des Grundwasserflurbstandes
Start	Beginn der Zeitreihe
End	Ende der Zeitreihe
n_points	Anzahl der Messwerte
duration_m	Dauer der Zeitreihe in Monaten
duration_y	Dauer der Zeitreihe in Jahren

avg_p_y	Durchschnittliche Anzahl an Messungen pro Jahr
points_spring_summer	Anzahl der Messungen in Frühjahr-Sommer
points_autumn_winter	Anzahl der Messungen in Herbst-Winter

## 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).

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

- ▶ Niedersachsen (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN))
- ▶ Nordrhein-Westfalen (Landesamt für Natur, Umwelt und Klima NRW)
- ▶ Rheinland-Pfalz (Landesamt für Umwelt Rheinland-Pfalz)
- ▶ Schleswig-Holstein (Landesamt für Umwelt des Landes Schleswig-Holstein (LfU))
- ▶ Saarland (Landesamt für Umwelt und Arbeitsschutz (LUA))
- ▶ 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)

## Literatur

- 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>.
- BGR (2019), Mittlere jährliche Grundwasserneubildung von Deutschland 1:1.000.000 (GWN1000) (WMS). Bundesanstalt für Geowissenschaften und Rohstoffe (Hrsg) Hannover.
- 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>.
- Cooke, A.-K.; Dorau, K.; Erkens, E.; Joger, F.; Willkommen, S.; Broda, S. (2025), Grundwasservulnerabilitätsbewertung im ZulassungsverFAhren von PflanzenschutzMitteln (FARM). - UBA Texte XX/2025. . <XXX>(accessed XX.XX.XXXX)
- 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.
- 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>.