

Nitrat-Monitoring 4.0

Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser (NiMo 4.0)

Kurzbericht

Fraunhofer IOSB

Nitrat im Grundwasser stellt weltweit unter anderem für die Trinkwasserversorgung ein großes Problem dar. Die Verteilung von Nitrat im Grundwasser ist dabei das Ergebnis eines komplexen Zusammenspiels vieler Einflussfaktoren, welches sich mit herkömmlichen Modellen für große Gebiete aufgrund der hohen Komplexität der Domäne nur schwer modellieren lässt. KI-Anwendungen, insbesondere Neuronale Netze bzw. Deep Learning Verfahren, lassen als datenbasierte Modelle, die komplexe Zusammenhänge aus einer großen Datenmenge extrahieren und übertragen können, hier einen deutlichen Mehrwert bei der zeitlich-räumlichen Vorhersage von Nitratwerten erwarten. Im Projekt NiMo 4.0 wurde daher ein übergreifendes System entwickelt werden, welches KI Verfahren mit Methoden der Umweltinformatik und speziell der Wasserdomäne kombiniert. Um verschiedene Datenquellen und KI-Verfahren auch im operativen Betrieb effektiv nutzen zu können, wurden neue Entwicklungen der intelligenten Sensordatenübertragung und -verarbeitung auf Basis des neuen offenen Standards SensorThingsAPI des Open Geospatial Consortiums (OGC) praxistauglich gemacht und in verschiedenen Demonstratoren von Endanwenderplattformen eingesetzt.

Aufgabe des Fraunhofer IOSB war die Spezifikation der Systemarchitektur auf Basis offener Standards gemeinsam mit Projektleiter disy und die Umsetzung dieser für ein intelligentes Datenmanagement. Hierfür wurde die open source Implementierung FROST® der SensorThings API der OGC eingesetzt und durch Fraunhofer IOSB weiterentwickelt. In NiMo 4.0 wurde zunächst ein FROST-Server® mit Nitratdaten der Pilotregionen in Baden-Württemberg und Niedersachsen aufgesetzt¹ Die Daten standen damit direkt nutzbar für die Entwicklung, das Training und die Ausführung der verschiedenen in NiMo 4.0 zu erforschenden ML-Algorithmen durch die weiteren Partner AGW, TZW und die Abteilung UWR des IOSB direkt über die STA-Schnittstelle zur Verfügung. Da die ML-Methoden in NiMo 4.0 in Python implementiert wurden wurde für den FROST-Server® eine neue Client-Bibliothek der Python-Client entwickelt, welcher nun ebenfalls open source zur Verfügung steht^{2,3}. Im Verlauf des Projektes haben weitere Landesumweltämter⁴ Interesse an den Entwicklungen gezeigt und ihrerseits Daten zur Verfügung gestellt. Damit konnte ein Datenserver mit einem länderübergreifenden Datenschatz von über 6 Millionen Nitratmesswerten und mit Nitrat in Beziehung stehenden chemischer Parametern wie Grundwasserstand, Leitfähigkeit, Temperatur, pH-Wert, Sauerstoff und Trübung aufgebaut werden. Im Rahmen von NiMo 4.0 wurde dabei auch die Echtzeitanbindung von Datenloggern für Nitratmesswerte und die kontinuierliche Integration der Daten in den FROST-Server® erfolgreich getestet.

¹ Hierfür wurden Daten von der LUBW, LW Baden-Württemberg und die WAZ in Niedersachsen bereitgestellt.

² [FROST®-Server - Open-Source-Implementierung der OGC SensorThings API - Fraunhofer IOSB](#)

³ [Fraunhofer IOSB · GitHub](#)

⁴ Vielen Dank an Rheinland-Pfalz (LfU), Thüringen (TLUBN), Niedersachsen (NLWKN) und Brandenburg

Neben den Funktionen für den Datenzugriff bietet die SensorThingsAPI eine Tasking Erweiterung an, mit welcher die Ausführung von Tasks (Aufgaben) über die Standardschnittstelle des STA möglich ist. Diese Schnittstelle wurde in NiMo 4.0 erstmalig für die Ausführung von ML-Methoden getestet. Hierfür wurde die Implementierung PERMA weiterentwickelt, mit welcher es möglich ist die ML-Modelle in einem Docker Container zu kapseln und über die STA zur Ausführung zu bringen.

Das Fraunhofer IOSB hat zwei weitere Aufgaben erfolgreich umgesetzt. Zum einen wurde die Verwendbarkeit von Convolutional Neural Networks zur Regionalisierung von Nitrat untersucht. Die Machine-Learning-Modelle wurden in Python programmiert und in einer Fallstudie in Baden-Württemberg getestet. Zum anderen wurden neue Methoden auf Basis von Deep Learning (DL) entwickelt und getestet, um Anomalien und Ereignisse in multivariaten raumzeitlichen Informationen aus Grundwasserdaten zu identifizieren. Eine Fallstudie wurde für Niedersachsen und Thüringen durchgeführt.

Die Entwicklungen wurden schlussendlich für die Umsetzung der anwenderspezifischen Entscheidungsunterstützung herangezogen. Auf Basis von WebGenesis®, einem System zur Vernetzung von Daten und zur Entscheidungsunterstützung wurde von Fraunhofer IOSB beispielhaft eine NiMo-Endanwenderplattform aufgebaut. Über die Plattform sind alle Daten des FROST-Servers via Karten und Diagrammdarstellungen direkt für den Endanwender zugreifbar. Die Plattform bietet eine einfach verständliche Nutzeroberfläche für die Ansteuerung der in NiMo 4.0 entwickelten ML-Algorithmen, sowie Möglichkeiten zur Visualisierung der Prognoseergebnisse der Algorithmen. Zusätzlich wurde in NiMo 4.0 eine bewusst einfache Ontologie für die Beschreibung der ML-Algorithmen entwickelt. Diese wurde in WebGenesis® integriert zum Aufbau einer Wissensbasis, welche die NiMo 4.0 ML-Algorithmen beschreibt und damit das Verständnis für die ML-Methoden und deren Anwendungsmöglichkeiten durch den Endnutzer erhöht.

The screenshot shows the 'Event Detection System' page in the NiMo application. It features a navigation menu at the top with options like 'Einführung', 'Ontologie', and 'Ergebnisse'. The main content area is titled 'Event Detection System' and contains a list of detectors: ForecastDetector, ReconstructionDetector, ClusteringDetector, UnivariateOutlierDetector, SignalQualityChangeDetector, and ViolationDetector. Below this, there are sections for 'Anwendung des EDS in NiMo' and 'Inhalte', which lists various data files like 'Bld1 maed gnubbs.png' and 'Bld3 reconstruction.png'.

Abbildung 1: Beschreibung eines ML-Algorithmus für den Endnutzer in der Ontologie

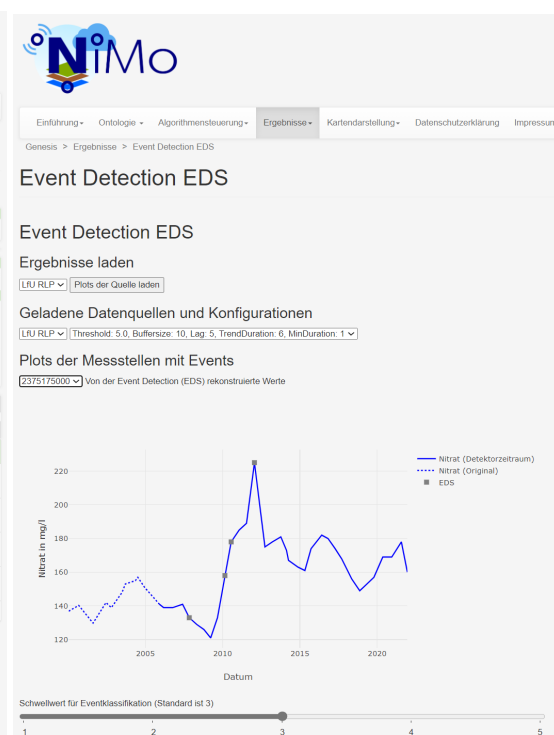


Abbildung 2: Ergebnisvisualisierung des Event Detection Modells

Nitrat-Monitoring 4.0

Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser (NiMo 4.0)

Inhaltlicher Schlussbericht

Fraunhofer IOSB

Nitrat im Grundwasser stellt weltweit unter anderem für die Trinkwasserversorgung ein großes Problem dar. Die Verteilung von Nitrat im Grundwasser ist dabei das Ergebnis eines komplexen Zusammenspiels vieler Einflussfaktoren, welches sich mit herkömmlichen Modellen für große Gebiete aufgrund der hohen Komplexität der Domäne nur schwer modellieren lässt. KI-Anwendungen, insbesondere Neuronale Netze bzw. Deep Learning Verfahren, lassen als datenbasierte Modelle, die komplexe Zusammenhänge aus einer großen Datenmenge extrahieren und übertragen können, hier einen deutlichen Mehrwert bei der zeitlich-räumlichen Vorhersage von Nitratwerten erwarten. Im Projekt NiMo 4.0 wurde daher ein übergreifendes System entwickelt werden, welches KI Verfahren mit Methoden der Umweltinformatik und speziell der Wasserdomäne kombiniert. Um verschiedene Datenquellen und KI-Verfahren auch im operativen Betrieb effektiv nutzen zu können, wurden neue Entwicklungen der intelligenten Sensordatenübertragung und -verarbeitung auf Basis des neuen offenen Standards SensorThingsAPI des Open Geospatial Consortiums (OGC) praxistauglich gemacht und in verschiedenen Demonstratoren von Endanwenderplattformen eingesetzt.

Einführung

Mit NiMo 4.0 sollte der Transfer von innovativen Lösungsansätzen von KI-Anwendungen im Grundwasser-Bereich von der universitären und institutionellen Forschung in die industrielle, anwendungsnahe Forschung und Praxis vorangetrieben werden. Thematische war die Entwicklung verschiedener KI-Ansätze zur Erforschung des Nitrats im Grundwasser maßgebend. Hierfür wurden neue KI-Methoden für die räumliche Ausbreitung von Nitrat im Grundwasser (Regionalisierung), für die Grundwassermessnetzoptimierung und für die Anomaliedetektion in Nitratzeitreihendaten entwickelt.

Hauptziel des IOSB im Projekt war die Realisierung des Datenmanagements als notwendige Grundlage für die Erforschung der KI-Methoden mittels aktueller offene Geodienste-Standards der OGC. Diese sollten für ihre Eignung bei der Integration von KI-Algorithmen untersucht und weiterentwickelt werden. Die im Gegensatz zu anderen OGC Standards neue leichtgewichtige SensorThings API, mit ihrer Erweiterung Task Core als Möglichkeit zur KI-Integration bietet das Potential, als einfach zu handhabende Schnittstelle für die Datenintegration zu dienen, weil sie sowohl in bestehende Systeme als auch direkt gemessene Sensordaten oder Ergebnisse integrieren kann. Die dabei entstehende Kombination des harmonisierten Datenzugriffs über offene FuE-Verbundvorhaben und die Nutzung dieser Methoden für die KI-Integration eröffnet auch aufgrund der technischen Übertragbarkeit bis zu Projektbeginn nicht realisierte Möglichkeiten der systemübergreifenden Datenauswertung von jahrzehntelang gesammelten Daten in Kombination mit neuen Messdaten und Algorithmen.

Eine weitere Hauptaufgabe des IOSB war die Nutzung der Datenmanagementkomponenten und die Erforschung neuer Methoden zur anwenderspezifischen Entscheidungsunterstützung für die

Realisierung von Demonstratoren für Endnutzer. Ziel war es Möglichkeiten zu schaffen Endanwender die Nutzung der KI-Komponenten und Analyse der Ergebnisse zu ermöglichen.

Beitrag zur Anforderungsanalyse und Systemarchitektur

Zu Beginn des Projektes wurde gemeinsam mit den Projektpartnern in Arbeitstreffen die Daten- und Systemanforderungen für die Erforschung der KI-Methoden definiert. Dabei wurden die folgenden Fragestellungen berücksichtigt:

- Was ist das Ziel der KI-Methode?
- Anhand welches Use Cases wollen wir das Ziel der KI-Methode in NiMo verfolgen?
- Welche Daten sind dafür erforderlich?
- Welches sind die Input-Parameter für eine KI-Methode?
- Welches Ergebnis wird erwartet?

Auf Basis dieser Informationen wurde gemeinsam mit dem Partner disy eine Systemarchitektur abgeleitet, welche in den nachfolgenden Arbeitspaketen im Rahmen des Projektes beispielhaft realisiert wurde.

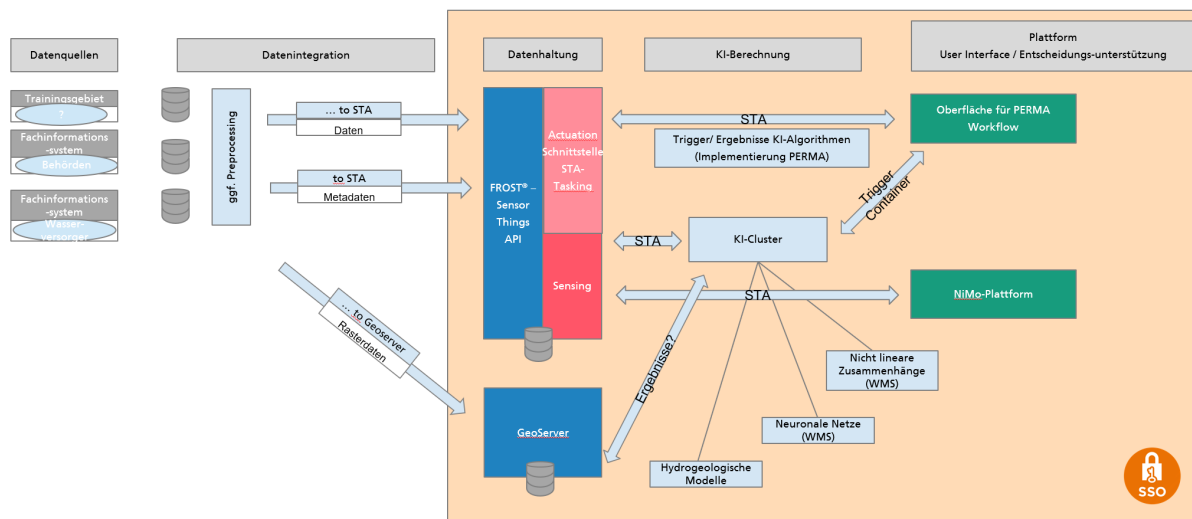


Abbildung 1: Systemarchitektur von NiMo 4.0

Intelligentes Datenmanagement für das Grundwassermonitoring

Die Grundlage für die Daten- und KI-Integration in NiMo 4.0 bildet der offene und leichtgewichtige Geodatenstandard SensorThings API des Open Geospatial Consortiums¹. Dabei wird für die Bereitstellung der messreihenbasierten Daten auf den ersten Teil (Part 1: Sensing) des Standards gesetzt. Zur Einbindung der KI-Algorithmen, deren Konfiguration und Ausführung wird auf Teil Zwei (Part 2: Tasking Core) zurückgegriffen.

Datenmodellierung mittels SensorThings API

Die OGC SensorThings API ist ein auf modernen Konzepten und Technologien aufgebauter Standard zur Verwaltung von Sensormess- und -metadaten aus dem IoT Kontext. Sie definiert eine JSON-basierte Datenrepräsentation, eine REST-Schnittstelle zur Datenabfrage und Manipulation sowie eine auf OData aufbauende Abfrage- und Filtersprache. Dies ermöglicht eine einfache und technologieunabhängige Bereitstellung von zeitreihenbasierten Daten. Dadurch hat die Verbreitung der SensorThings API für den Zugriff auf Sensordaten in den letzten Jahren zunehmend an Bedeutung gewonnen.

¹ [OGC SensorThings API - Open Geospatial Consortium](#)

Die OGC SensorThings API besteht aus den folgenden Komponenten:

- Einem Datenmodell, welches vorhandene Entitäten und deren Beziehungen definiert
- Einer HTTP basierte Schnittstelle, die das Abfragen, Bearbeiten und Löschen von Daten erlaubt
- Mehreren Erweiterungen, wie beispielsweise einer MQTT-Schnittstelle, welche die Umsetzung ereignisbasierter Anwendungsfälle erlaubt.

Ziel in NiMo 4.0 war die Messdaten verschiedenster Messstellen und unterschiedlicher chemischer Parameter, bereitgestellt in verschiedenen Datenmodellen und Datenformaten einheitlich über das Datenmodell der SensorThings API zu verwalten und in der definierten Systemarchitektur zur Verfügung zu stellen.

Hierfür wurde das folgende Mapping von durch Anwendungspartner bereitgestellten Daten auf das STA Datenmodell definiert. Es wurde als grundlegende Vorlage für die Integration von Gewässerqualitätsdaten in NiMo 4.0 verwendet.

| Entität der SensorThings API | Verschiedene Datenquellen |
|------------------------------|---|
| Thing | Messstelle |
| Location | Koordinaten der Messstelle |
| ObservedProperty | Chemischer Parameter |
| Sensor | Ggf. Informationen über das Messgerät oder den Sensor |
| Datastream | Informationen über die Messreihe |
| Observation | Messwert |
| FeatureOfInterest | Probenahme |

Als Implementierung der SensorThings API wurde im NiMo 4.0 Porjekt die open source Referenz-Implementierung FROST-Server® verwendet, die am Fraunhofer IOSB entwickelt wurde.²

Datenmanagement für NiMo 4.0 in FROST

In NiMo 4.0 wurde zunächst für das Datenmanagement mit Hilfe des FROST-Server® eine Instanz der SensorThings API mit gemessenen Nitratdaten der Pilotregionen in Baden-Württemberg und Niedersachsen aufgesetzt. Im Fokus standen Nitratdaten, aber auch weitere Wasserqualitätsdaten, die in Zusammenhang mit der Nitratproblematik stehen. Neben den bereits bei Projektantrag angefragten Datenlieferanten WAZ, Zweckverband Landeswasserversorgung und LUBW haben im Verlauf des Projektes weitere Länder Interesse am Projekt gezeigt und Daten für die Forschung zur Verfügung gestellt. Diese wurden im Verlauf des Projektes kontinuierlich in den FROST®-Server mit den Basisdaten integriert. Folgende Länder und Wasserversorger haben mit Daten zum Erfolg des NiMo-Projekts beigetragen:

- Umweltämter:
 - Landesamt für Umwelt Brandenburg
 - Thüringer Landesamt für Umwelt, Bergbau und Naturschutz
 - Drei Echtzeitdatenlogger von SEBA Hydrometrie GmbH & Co. KG
 - Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz
 - Landesamt für Umwelt Rheinland-Pfalz
 - Landesanstalt für Umwelt Baden-Württemberg
- Wasserversorger
 - Zweckverband Landeswasserversorgung
 - Wasser- und Abwasser-Zweckverband Niedergrafschaft

² [FROST®-Server - Open-Source-Implementierung der OGC SensorThings API - Fraunhofer IOSB](#)

- Grundwasserdatenbank Wasserversorgung

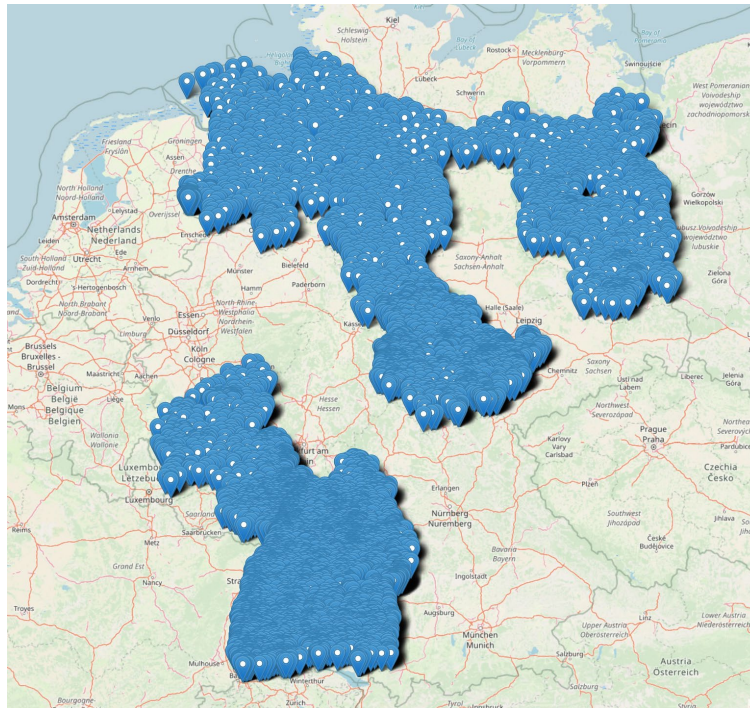


Abbildung 2: Messstellenübersicht der Gewässerqualitätsdaten des FROST®-Servers

Die Gewässerqualitätsdaten wurden auf verschiedene Arten bereitgestellt. Der FROST®-Server bietet sowohl Mittel zur Integration von Daten, bereitgestellt als CSV-Dateien, sowie Programmbibliotheken zur Integration von Daten über die STA-Schnittstelle. Beides kam im NiMo 4.0 Projekt zum Einsatz. Neben den klassischen Messdaten wurden neu auch die Daten von drei Datenloggern aus Thüringen kontinuierlich und in Echtzeit in den FROST-Server® übertragen. Dies geschah über die MQTT-Schnittstelle.

Letztendlich konnte damit in NiMo erstmalig eine Forschungsinstanz des FROST-Servers mit nitratrelevanten Daten (Nitrat, Grundwasserstand, Leitfähigkeit, Temperatur, pH-Wert, Sauerstoff, Trübung und weitere) aus vielen Teilen Deutschlands mit über 7000 Messstellen und ca. 6.5 Millionen Messwerten von 1981 bis heute aufgebaut werden.

Auf die integrierten Messdaten konnte dann jederzeit mittels REST-Abfragen über die http-Schnittstelle der SensorThings API zugegriffen werden. Dadurch standen die aktuellsten Daten direkt sowohl den KI-Komponenten als auch Endanwenderplattformen zur Verfügung.

Da im Zuge von KI-Algorithmen häufig Python als Programmiersprache zum Einsatz kommt, wurde im Rahmen des Projekts mit dem Python Client ein open source Tool geschaffen, was die Interaktion mit einem FROST-Server® erleichtert, etwa um Daten abzurufen oder Ergebnisse zu schreiben³.

Integration von KI-Modellen

Neben der SensorThings API Part 1: Sensing, existiert ein weiterer Standard, die SensorThings API Part 2: Tasking Core. Dieser Teil des Standards verfügt über die interessante Möglichkeit mittels der leichtgewichtigen Schnittstelle nicht nur Daten abzufragen, sondern auch Berechnungsprozesse anzustoßen. Das Datenmodell des ersten Teils wird um weitere Entitäten ergänzt. Dadurch können mit Hilfe der SensorThings API nicht nur Sensor Mess- und Metadaten,

³ <https://github.com/FraunhoferIOSB/FROST-Python-Client>

sondern auch Aktuatoren modelliert werden. Dies bietet die Grundlage für die Integration der in NiMo entwickelten KI-Algorithmen. Die Fähigkeiten des Aktors werden als TaskingCapability dargestellt. Diese wiederum sind einem Thing zugeordnet, was den Übergang in das Datenmodell des ersten Teils darstellt. Während eine TaskingCapability die Fähigkeiten beschreibt, werden Aktionen des Aktors mittels eines Tasks umgesetzt. Akzeptierte Konfigurationsparameter einer TaskingCapability werden über das Attribut taskingParameters definiert. Die konkreten Werte zur Ausführung werden im Task über das gleichnamige Attribut abgelegt. Aufrufe an die Softwarekomponente können somit als Task und Aufrufparameter als taskingParameter dargestellt werden. Betrachtet man mit diesem Hintergrund einen KI-Algorithmus, so kann dieser als virtueller Aktor aufgefasst und das Anstoßen einer Berechnung als Task umgesetzt werden. Dieses Konzept wurde in NiMo 4.0 für die Integration der KI-Algorithmen umgesetzt. Das Verfahren ist in der Komponente PERMA[®] prototypisch implementiert. Die KI-Algorithmen können in PERMA[®] in Docker-Containern gekapselt werden und die Berechnung der jeweiligen KI-Komponenten im Container werden über die STA-Schnittstelle mittels Tasks gestartet. Wenn ein entsprechender Task im FROST-Server erstellt wird, stellt PERMA[®] sicher, dass der entsprechende (KI-)Algorithmus auf einer Container-Plattform wie Kubernetes wird. Vortrainierte Modelle können aus einem Cloud-Speicher in den Container geladen werden. Die Ergebnisse der Berechnungen können dann wieder im FROST[®]-Server gespeichert werden. Im Rahmen von NiMo wurde PERMA[®] mit der PyPerma-Komponente erweitert, welche Support für die Integration von in Python implementierten KI-Algorithmen bietet.

Zeitreihenbasierte Prognoseergebnisse werden im NiMo Datenmanagement ebenfalls im Datenmodell der STA abgelegt. Um Mess- von Prognosedaten klar zu trennen, wurde eine separate FROST-Server[®] Instanz zum Speichern von Resultaten verwendet. Die Zugriffsrechte auf Roh- und Ergebnisdaten können unterschiedlich gehandhabt werden. Für die Ablage von Prognoseergebnissen anderer Formate wurde in NiMo das open source Produkt GeoServer verwendet, welches weitere OGC-Standards unterstützt.

Abbildung 3 zeigt das auf Basis der SensorThings API aufgebaute Datenmanagement mit Hilfe des FROST[®]-Servers. Damit standen die grundlegenden Komponenten um die Datenverarbeitungskette im Zusammenspiel mit den von den KI-Experten entwickelten KI-Modellen in den NiMo Themen Nitratregionalisierung, Messnetzoptimierung und Anomaliedetektion zu testen.

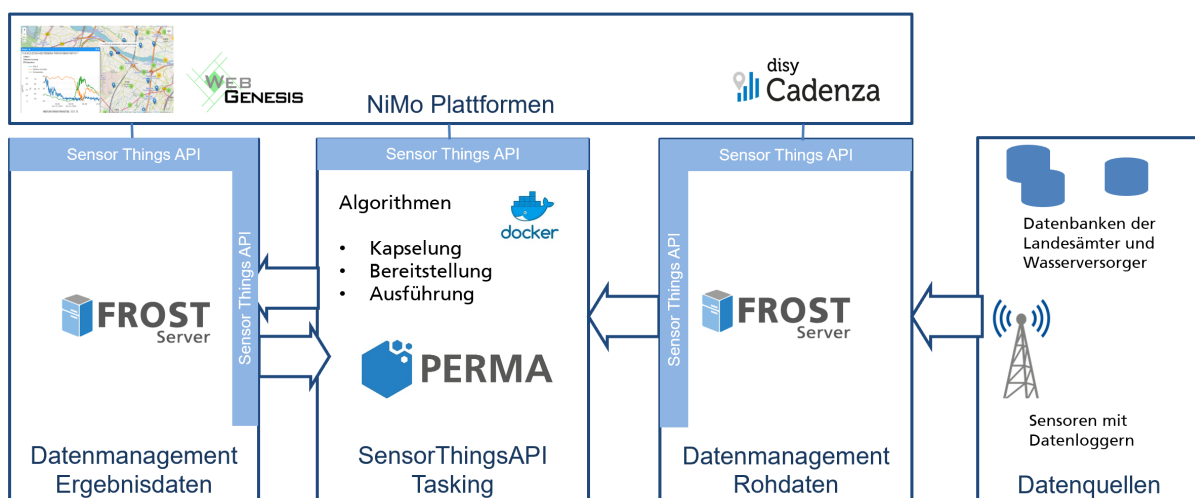


Abbildung 3: Datenmanagement mittels SensorThings API

Beitrag zum räumlichen Maschinellen Lernen für Grundwasserqualitätsmonitoring

Um die Regionalisierung von Nitrat im Grundwasser zu untersuchen, wurden zwei grundlegende Prinzipien untersucht (siehe Abb. 4). Die Methoden umfassen baumbasierte Methoden, die auf einer punktwisen Betrachtung basieren und durch Extra Trees Regressor repräsentiert werden, sowie bildbasierte Methoden (Kontextbetrachtung), die durch Convolutional Neural Networks repräsentiert werden. Die Machine-Learning-Modelle wurden in Python implementiert und an einer Fallstudie in Baden-Württemberg getestet.

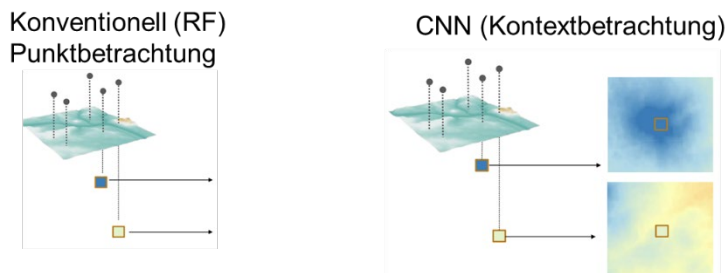


Abbildung 4: Betrachtung des räumlichen Kontextes in der Nähe einer Nitratmessstelle für eine bestimmte Anzahl von erklärenden Rastern

Der erste Ansatz, Extra Trees Regressor, ist ein auf Baumstrukturen basierender Ansatz, der eine Kombination von Regressionsbäumen verwendet, um die Regionalisierung von Nitraten im Grundwasser vorherzusagen. Bei diesem Ansatz werden die Daten rekursiv in immer kleinere Untergruppen aufgeteilt, wobei jede Partition durch einen Baumzweig repräsentiert wird. Die endgültige Vorhersage erfolgt durch die Aggregation der Vorhersagen aller Bäume im Modell. Dieser Ansatz eignet sich gut, um komplexe, nichtlineare Beziehungen zwischen den Eingangsvariablen und der Regionalisierung von Nitraten im Grundwasser zu erfassen.

Der zweite Ansatz, Convolutional Neural Networks (CNNs), ist ein bildbasierter Ansatz, der häufig bei Aufgaben der Computer Vision verwendet wird, aber auch auf geografische Daten angewendet werden kann. CNNs verwenden mehrere Schichten von Faltungsfilttern, um Merkmale aus den Eingangsdaten zu extrahieren, die dann zur Vorhersage verwendet werden. Im Kontext der Regionalisierung von Nitrat im Grundwasser können CNNs verwendet werden, um Rasterbilder als Eingangsdaten zu analysieren, Muster und Merkmale in diesen zu identifizieren, und damit auf hohe oder niedrige Nitratwerte im Grundwasser zu schließen.

Diese beiden Methoden wurden für die Regionalisierung von Nitrat im Grundwasser verglichen, um den effektiveren Ansatz für diese spezifische Aufgabe zu bestimmen. Darüber hinaus möchten wir, durch die Erkundung von baum- und bildbasierten Methoden ein besseres Verständnis der zugrunde liegenden Muster und Beziehungen in den Daten gewinnen, und so mit zukünftiger Forschung und Entscheidungsprozessen im Zusammenhang mit der Grundwasserverwaltung und dem Umweltschutz informieren.

Die Struktur der baumbasierten Methode erlaubt nur die Eingabe von Zeitreihendaten ohne räumliche Informationen. Daher wurden für die Untersuchung die Eingabedaten auf zwei Arten erweitert: (1) mit konstruierten Merkmalen, die räumliche Informationen wie die Entfernungen von den Ecken oder dem Zentrum des Forschungsgebiets enthalten, und (2) mit Deepkriging, eine Methode, die Deep Learning mit Kriging, einer geostatistischen Interpolations-Technik, kombiniert. Dadurch werden räumliche Information in die Eingabedaten integriert, was die Genauigkeit und Leistung der baumbasierten Methode verbessert.

Der Nachteil der Deepkriging-Methode ist, dass sie die Anzahl der Hyperparameter des zu schätzenden Modells erhöht. Darüber hinaus bezieht Kriging die Koordinaten der Region mit ein,

was die Ergebnisse für die lokale Region erheblich verbessert, aber die Verallgemeinerung des Deep-Learning-Modells verringern könnte.

Alle Methoden wurden in Python implementiert und können auf einer CPU ausgeführt werden. Dadurch kann die Implementierung auf einer großen Auswahl von Geräten und Plattformen funktionieren, was die Zugänglichkeit und Anwendbarkeit der Methoden erhöht.

Außerdem ist es wichtig, dass Umweltmanager, Wasserbehörden und Wasserversorger die Zusammenhänge zwischen verschiedenen Faktoren und den Vorhersagen von Modellen verstehen, um die Qualität des Grundwassers zu verbessern. Zu diesem Zweck wurden Methoden des Feature-Importance-Learning und der partiellen Abhängigkeitsanalyse implementiert, um ein besseres Verständnis der Daten und Modellvorhersagen zu ermöglichen. Das Feature-Importance-Learning hilft dabei, herauszufinden, welche Faktoren den größten Einfluss auf die Vorhersagen des Modells haben, während die partielle Abhängigkeitsanalyse zeigt, wie Veränderungen bestimmter Faktoren die Vorhersagen des Modells beeinflussen. Durch die Erkenntnisse aus diesen Analysen können Stakeholder Maßnahmen zur Verbesserung der Grundwasserqualität auf der Grundlage eines besseren Verständnisses der Einflussfaktoren getroffen werden. Diese Informationen sind entscheidend für fundierte Entscheidungen und wirksame Maßnahmen zum Schutz und zur Bewirtschaftung von Wasserressourcen.

Darüber hinaus wurden in früheren Studien große Unsicherheiten bei der Vorhersage von Nitrat festgestellt. Aus diesem Grund wurden die Modelle erweitert, um Unsicherheiten mithilfe von Vorhersageintervallen (PIs) zu quantifizieren, die aus der Methode Bootstrapping abgeleitet wurden. Das Wissen über Unsicherheiten hilft potentiellen Anwendern das Risiko zu reduzieren und zuverlässiger zu planen.

Es konnte festgestellt werden, dass die baum- und bildbasierten Methoden eine bessere Regionalisierung von Nitrat im Grundwasser ermöglichen, verglichen mit den zuvor angewandten Methoden wie z.B. SIMCOP-BW, wie in Tabelle 1 dargestellt.

Tabelle 1: Auswertung der Modelle für eine Fallstudie in Baden-Württemberg

| Metrik | Best RF | SIMCOP-BW | Best 2DCNN | 2DCNN Ensemble Mittelwert |
|---|----------------|------------------|-------------------|----------------------------------|
| Mittelwert absolute Fehler – MAE(std) (mg/l) | 9.55 | 13.94 | 10.836 (0.369) | 11.098 (0.300) |
| Mittelwert R2-Score | 0.33 | | 0.297 | 0.296 |

Die Ergebnisse der Modelle (Extra Tree Regressor (RF) und Varianten von CNNs) zur Regionalisierung des Nitratgehalts im Grundwasser in Baden-Württemberg zeigen, dass es regionale Unterschiede in der Nitratbelastung gibt, wie die Abb.5 zeigt. In einigen Gebieten ist der Nitratgehalt im Grundwasser sehr hoch, während er in anderen Regionen geringer ist. Diese Unterschiede können auf verschiedene landwirtschaftliche Praktiken, geologische Bedingungen und hydrologische Prozesse durch Partial Dependence Plots (PDP) veranschaulicht werden. Die Modelle können dazu beitragen, die Ursachen für die Nitratbelastung zu identifizieren und Maßnahmen zur Reduzierung der Belastung zu entwickeln.

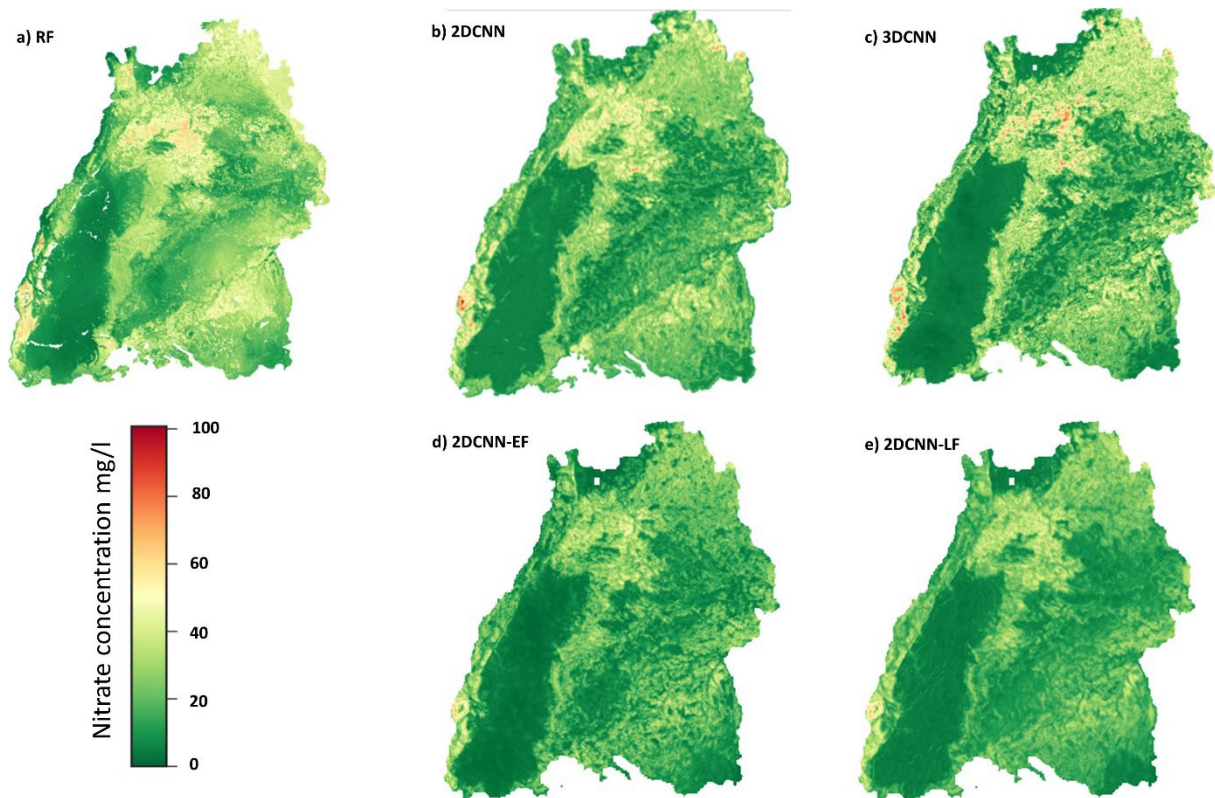


Abbildung 5: Regionalisierungsergebnis für Baden-Württemberg durch verschiedene Methoden

Weiterführende Beschreibungen der Methoden zur Regionalisierung von Nitrat im Grundwasser wurden in Karimanzira et al., 2023 a veröffentlicht.

Beitrag zur intelligenten Messwertüberwachung

Die Aufgabe des IOSB bestand darin, neue Methoden auf Basis von Deep Learning (DL) zu entwickeln, um multivariate raumzeitliche Informationen bei der Erkennung anomaler Ereignisse zu nutzen. Anomale Ereignisse sind sehr selten, wodurch die Beschaffung gelabelter Anomaliedatensätze herausfordernd ist. Daher war es sinnvoll, einen unüberwachten Ansatz für die Erkennung von Sensoranomalien und Ereignissen zu verwenden, wobei gelabelte Daten lediglich zur Festlegung von Schwellenwerten für die Entscheidung verwendet werden.

Es wurden zwei Methoden entwickelt. Die erste Methode besteht aus einem Long Short Term Memory (LSTM)-Encoder, einem LSTM-Decoder und einem LSTM-Predictor für die zeitliche Anomalieerkennung. Ein Deep Neural Network (DNN)-basierter Klassifizierer wird verwendet, um die codierte und trainierte latente Darstellung ihrer Herkunftsregion zuzuordnen, und damit einen zeitlichen und räumlichen Anomaliedetektor zu bilden.

Die zweite Methode basiert auf einem CNN-Encoder und einem LSTM-Decoder, um sowohl räumliche als auch zeitliche Merkmale zu erfassen. Die Encoder-Komponente kann entweder ein 3D-CNN oder ein Multikanal-CNN verwenden, um komplexe räumliche Abhängigkeiten zu erfassen. Der 3D-Tensor-Eingang für den Encoder wird gebildet, indem die Daten der nächsten räumlichen Nachbarn jedes Datenpunkts gestapelt werden.

Beide Modelle erkennen Anomalien in einem dreistufigen Prozess. Zuerst lernen die Modelle, basierend auf den Mustern und Merkmalen im Trainingsdatensatz, zwischen normalen und anomalen Datenpunkten zu unterscheiden. Nachdem die Modelle trainiert wurden, können sie zur Erkennung von Anomalien in neuen, nicht gesehenen Daten verwendet werden. Das Modell wird auf die neuen Daten angewendet und weist jedem Datenpunkt eine Wahrscheinlichkeit zu,

die angibt, wie wahrscheinlich es ist, dass es sich um eine Anomalie handelt. Datenpunkte mit Wahrscheinlichkeiten über einem bestimmten Schwellenwert werden als Anomalien klassifiziert, während diejenigen unter dem Schwellenwert als normal klassifiziert werden. Der spezifische Schwellenwert zur Klassifizierung von Anomalien kann je nach den spezifischen Anforderungen der Anwendung und dem Kompromiss zwischen falsch positiven und falsch negativen Ergebnissen angepasst werden. Darüber hinaus können Nachverarbeitungstechniken wie Ensemble-Methoden oder die Kombination beider Modelle zur Verbesserung der Genauigkeit der Anomalieerkennung verwendet werden.

In der dritten Stufe wird der Anomaliewert zerlegt, um die Attribute zu identifizieren, die am meisten zur Anomalie beitragen. Dieser Zerlegungsprozess hilft dabei, die spezifischen Attribute oder Merkmale zu identifizieren, die für den Anomaliewert verantwortlich sind, und gibt Einblicke in die Ursachen der Anomalie. Durch die Identifizierung der am meisten beitragenden Attribute ist es möglich, sich auf diese spezifischen Aspekte für weitere Untersuchungen und potenzielle Korrekturmaßnahmen zu konzentrieren. Dies kann letztendlich zu einem besseren Verständnis der zugrunde liegenden Faktoren führen, die die Anomalie beeinflussen, und helfen, gezielte Strategien zur Bewältigung und Verhinderung ähnlicher Anomalien in der Zukunft zu entwickeln.

Nachdem eine Anomalie erkannt und klassifiziert wurde, kann eine weitere Untersuchung durchgeführt werden, um den Verursacher zu identifizieren. Dies kann beinhalten, dass Experten aus dem jeweiligen Fachgebiet die Daten analysieren, um die Ursache der Anomalie zu ermitteln. In einigen Fällen können fortgeschrittene Analysetechniken wie Kausalitätsanalyse, Zeitreihenanalyse oder Ursache-Wirkungs-Modelle eingesetzt werden, um den Verursacher der Anomalie zu identifizieren. Es ist wichtig zu beachten, dass die Identifizierung des Verursachers einer Anomalie oft nicht automatisiert werden kann und menschliche Expertise und Domänenwissen erfordert. Daher ist die Zusammenarbeit zwischen Anomalieerkennungsmodellen und menschlichen Analysten oft entscheidend, um Anomalien effektiv zu verstehen und darauf zu reagieren.

Durch KI-basierte Methoden können verschiedene Arten von Anomalien und Ereignissen in den Daten zur Grundwasserqualität identifiziert werden. Neben Punktanomalien (Ausreißer, die z.B. durch Defekte in den Sensoren entstehen) können die Modelle auch Kontextanomalien, wie räumliche und zeitliche Ungleichmäßigkeiten sowie saisonale Veränderungen in den Parametern erkennen. All diese Arten von Anomalien sind mit herkömmlichen Methoden sehr schwer zu identifizieren. Beide Methoden liefern ähnliche Ergebnisse, aber unterscheiden sich in der Komplexität für die Ereigniserkennung und erkennen verschiedene Arten von Anomalien (Punkt, räumlich, Kontext usw.). Nach der Trainingsphase zur Erfassung des normalen Systemverhaltens sind beide Methoden in der Lage, Anomalien mit einer sehr guten Genauigkeit zu erkennen, die zuvor noch nie gesehen wurden (Werte zwischen 88% und 96%). Zur Validierung der Genauigkeit und Effizienz der DL-basierten Methoden wurden sie mit einem modifizierten ST-DBSCAN-Algorithmus verglichen. Die Ergebnisse zeigen die Überlegenheit der DL-basierten Methoden.

Die Ergebnisse können in Form eines Graphen, wie in Abbildung 6 dargestellt, veranschaulicht werden. Dadurch ist es dem Nutzer möglich, identifizierte multivariate Anomalien in Zusammenhang mit allen anderen Attributen (hier Nitrat und Wasserstand) zu erkennen.

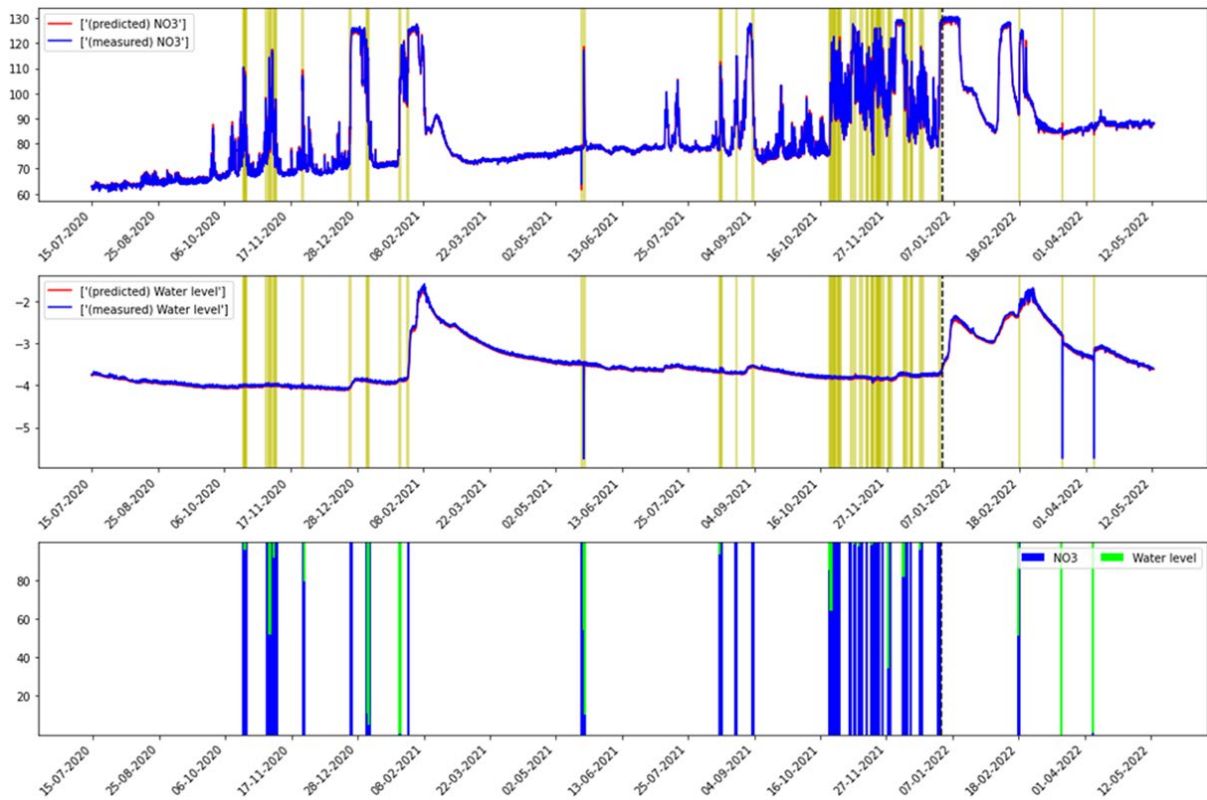


Abbildung 6: Grafische Visualisierung von Anomalien einer Messstelle und die prozentuale Beteiligung der Attribute zu Anomalien

Weiterführende Beschreibungen der Methoden zur **intelligenten Messwertüberwachung von Grundwassernetzen** wurden in Karimanzira et al., 2023 b veröffentlicht.

Anwenderspezifische Entscheidungsunterstützung

Die von KI-Experten entwickelten KI-Modelle wurden mit dem im Abschnitt „Datenmanagement für das Grundwassermonitoring“ erforschten Komponenten in die NiMo Systemarchitektur integriert. Das damit realisierte Datenmanagement auf Basis standardisierter offen definierter Schnittstellen ermöglicht den flexiblen Zugriff sowohl auf die zu Grund liegenden Basisdaten, als auch auf die Ergebnisdaten der in NiMo entwickelten KI-Prognosemodelle. Aufgabe der anwenderspezifischen Entscheidungsunterstützung ist die Entwicklung von Methoden für die Aufbereitung dieser Daten für Endanwender. Aufgabe des IOSB war die Erforschung und der Test dieser Methoden in einer beispielhaften Endanwenderplattform für Anwender, welche keine grundlegenden KI-Kenntnisse besitzen müssen.

Endanwenderplattform WebGenesis

Eine der zwei unterschiedlichen in NiMo 4.0 erforschten Endanwenderplattformen basiert auf WebGenesis®. WebGenesis® ist ein System zur Vernetzung einer Vielzahl von Daten zur Unterstützung von Entscheidungsprozessen. Es unterstützt auch semantische Web-Technologien. Dabei setzt WebGenesis auf offene Standards wie die Web Ontology Language OWL/RDF.⁴ Weiterhin bietet WebGenesis mit dem integrierten Rollenkonzept die Möglichkeit Anwendungssichten und Daten nur bestimmten Nutzergruppen freizugeben. Damit können Endanwendergruppen mit unterschiedlichem KI-Erfahrungshintergrund an sie angepasste Funktionalitäten freigegeben werden.

⁴ [WebGenesis - Fraunhofer IOSB](#)

Visualisierungsoptionen für Daten der SensorThings API

Eine Hauptaufgabe für eine NiMo Endanwenderplattform ist die Visualisierung von über die SensorThings API bereitgestellten Daten für den Endanwender.

In der Endanwender Plattform auf Basis von WebGenesis® wurden hierfür verschiedene Möglichkeiten realisiert. In einer Übersichtskarte sind Messstellen dargestellt, deren Messreihen chemischer Parameter für die Visualisierung in Zeitreihengrafiken flexibel ausgewählt und kombiniert werden können, siehe Abb. 7.

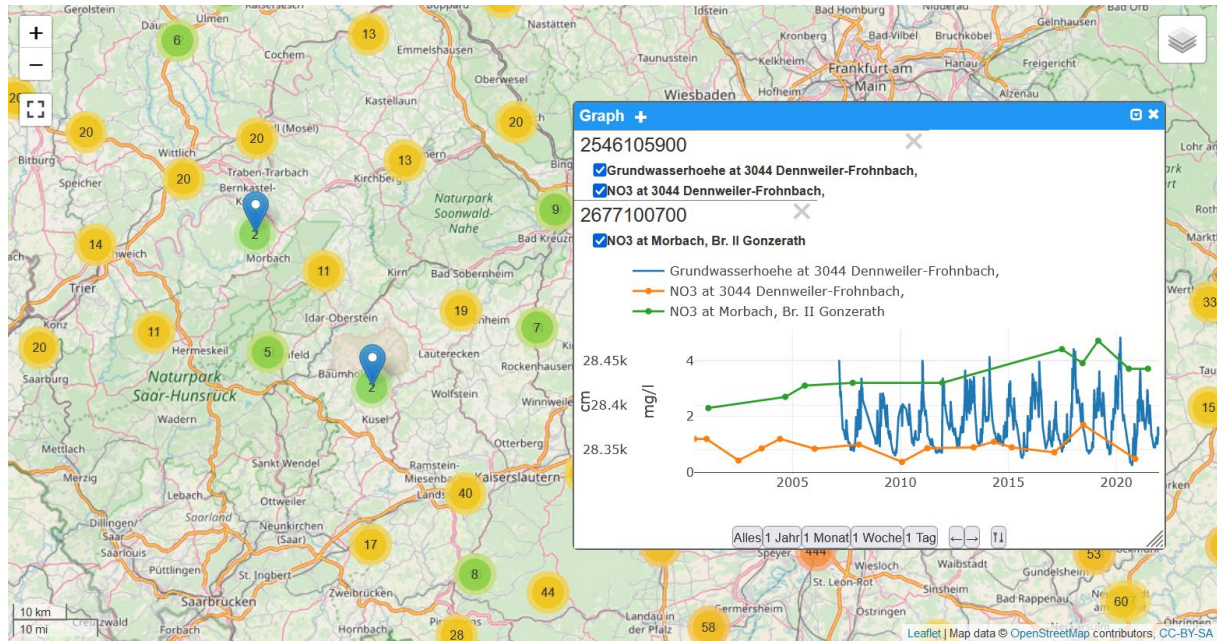


Abbildung 7: Kombinierte Visualisierung von Messreihen verschiedener Messstellen

Zusätzlich ist es möglich mit Hilfe eines Grafana Plugins die Daten des FROST®-Servers ohne großen Aufwand in einem Grafik-Dashboard aufzubereiten und zur Verfügung zu stellen.

Semantische Beschreibung der KI-Methoden in der NiMo 4.0 Ontologie

Bei den verschiedenen Endanwender-Workshops, die im Rahmen von NiMo stattgefunden haben, hat sich gezeigt, dass die Vermittlung von Wissen über KI-Methoden und deren Einsatzmöglichkeit für einen potentiellen Einsatz in der Praxis essentiell ist.

Deswegen wurde in NiMo eine Ontologie entwickelt, welche generell für die Beschreibung von KI-Methoden geeignet ist. Ziel war mittels dieser Ontologie die in NiMo 4.0 entwickelten KI-Module zu beschreiben und damit das Verständnis von Endanwendern für diese Methoden zu erhöhen. Zu diesem Zweck wurde die Ontologie speziell einfach gehalten. Die folgenden Klassen wurden in die Ontologie aufgenommen:

- *Thema* beschreibt das Themenfeld der Domäne. Der Benutzer wählt das Thema entsprechend seiner aktuellen Aufgabe.
- *Verfahrensanwendung* (Algorithmus) ist die Klasse zur Beschreibung eines spezifischen KI-Algorithmus. Hier kann sich der Nutzer über die Eigenschaften und Einsatzmöglichkeiten eines speziellen Algorithmus informieren.
- *Konfiguration* ist eine Unterklasse von Algorithmus und kann erfolgreich getestete Konfigurationen des Algorithmus mit verfügbaren Daten enthalten. Nutzer welche in NiMo integrierte KI-Modelle selbst anwenden möchten, finden hier Beispielkonfigurationen.

- *KI-Methode* beschreibt die angewandte KI-Methode, die zur Entwicklung des Algorithmus verwendet wurde.
- *Datensatz* beschreibt die Daten, die als erweiterte Datengrundlage für die Entwicklung des Algorithmus Verwendung fanden.
- *Parameter* beschreibt die Konfigurationsparameter des Algorithmus. Diese Informationen müssen vom Benutzer in einer Konfiguration angegeben werden, die den Algorithmus auf neue Daten anwenden.
- *Ergebnis* sammelt verschiedene Ergebnisse, die der Algorithmus berechnet hat

In WebGenesis® ist es möglich vorhandene Ontologien direkt zu importieren. Eingabe- und Suchformulare können automatisch generiert und eine Navigationsstruktur erstellt werden. Die Struktur kann in WebGenesis® erstellt und bearbeitet werden. Mit diesen Funktionen kann ein komplexes wissensbasiertes Informationssystem in kurzer Zeit aufgebaut werden. Die in NiMo 4.0 entwickelte Ontologie wurde also in der Endanwenderplattform von WebGenesis® genutzt um einen Wissensbasis der NiMo KI Methoden Nitratregionalisierung, Messnetzoptimierung und Anomalieerkennung aufzubauen.

Abbildung 8 zeigt beispielhaft eine Instanz der Klasse Algorithmus. Endnutzer können sich in der Plattform direkt über die Eigenschaften des Algorithmus informieren.

Nitratregionalisierung Random Forest

Vorprozessierung

Vor der Modellierung erfolgt eine gründliche Vorverarbeitung der Daten. Dies enthält die Berechnung eines vierjährigen Mittelwerts für den Nitratgehalt von 2016 bis 2019, eine Ausreißerbereinigung, sowie die Aussortierung von Messstellen in tieferen Stockwerken, sofern relevante Informationen über die Stockwerkszuordnung vorliegen.

An den Nitratmessstellen dienen Rasterwerte aus zahlreichen Rastern mit Umweltinformationen als Prädiktoren. Die verwendeten Datensätze sind rechts verlinkt.

Die eigentliche Auswahl der relevantesten Prädiktoren erfolgt später durch Bayes'sche Optimierung. Diese Methode basiert auf einem probabilistischen Modell der Zielfunktion und automatisiert die Suche nach optimalen Parameterwerten, um die Modellleistung zu verbessern.

Die räumlichen Beziehungen zwischen den Messstellen werden durch eine Distanzmatrix dargestellt, die alle paarweisen euklidischen Entfernungen zwischen den Messstellen enthält. Aufgrund der umfangreichen Datenmenge wird eine Hauptkomponentenanalyse (PCA) durchgeführt, um nur die entscheidenden Hauptkomponenten zu berücksichtigen.

Trainingsprozess

Nach der Vorverarbeitung der Eingangsdaten beginnt der Trainingsprozess des Random Forest Modells. Das Modell nutzt die vorliegenden Trainingsdaten, d.h. die Hauptkomponenten der Distanzmatrix und die Rasterdaten mit Umweltinformationen als Prädiktoren, um den Zusammenhang mit den Nitratwerten als Zielwerten zu erlernen. Im Trainingsprozess wird gleichzeitig mittels k-fold Kreuzvalidierung eine Modellevaluierung vorgenommen, d.h. es wird berechnet, wie groß im Mittel der Fehler sein kann, der später in der eigentlichen Regionalisierung in Kauf genommen werden muss.

Regionalisierung

Nach Training und Evaluierung erfolgt die eigentliche Regionalisierung. Hierbei schätzt das trainierte Modell den Nitratgehalt an einem bestimmten Standort basierend auf den Prädiktoren an ausgewählten Datenpunkten mit dem Ergebnis einer räumliche Vorhersage des Nitratgehalts im Grundwasser für bisher nicht beprobte Bereiche.

Verwendet Datensatz

- Bodenübersichtskarte 1:1.000.000 (BÜK1000); 1.ethodenassoziationen
- Bodenübersichtskarte 1:200.000 (BÜK200); Bodentypen
- Corine Landcover 2018
- Digitales Höhenmodell von Deutschland (DHM)
- Gehalte organischer Substanz in Oberböden 1:1.000.000 (HUMUS1000 OB); Oberboden
- Geologische Karte von Deutschland 1:1.000.000 (GK1000)
- Hydrogeologische Raumgliederung von Deutschland 1.500.000 (HYRAUM), Teilraum
- Hydrogeologische Übersichtskarte von Deutschland 1:250.000 (HUK250)
- Normalized Difference Vegetation Index (NDVI)
- Preidl Landnutzungsdaten
- Slope (Hangneigung)
- Stream Distance (Abstand zu Fließgewässer)
- Thünen Landnutzung
- Topographic Wetness Index (TWI)

Abbildung 8: Nitratregionalisierung Random Forest. Neben der Beschreibung des Algorithmus sieht man auf der rechten Seite die Navigationslinks zu den Datensatzinstanzen, die vom Algorithmus verwendet wurden

Algorithmenintegration in WebGenesis

Neben der Ontologie bietet die WebGenesis®-Plattform benutzerfreundliche Möglichkeiten zur Interaktion mit den Algorithmen. Umsetzungsziel hierfür war die Steuerung der KI-Modelle mittels SensorThings API Tasks über die Implementierung PERMA®. Die für die Tasks definierten Attribute können vom Benutzer über die Oberfläche der Endanwenderplattform definiert werden. Diese werden dann über PERMA® an den FROST-Server® zur Ausführung der Algorithmen übermittelt. Verschiedene Parameter können vom Benutzer konfiguriert werden. Der Benutzer kann die spezifische Bedeutung jedes Parameters über die verlinkte Ontologieeinträge erfahren oder einfach einen Standardwert verwenden. Entsprechende Funktionsaufrufe sind in der Plattform im Menüeintrag Algorithmensteuerung gesammelt. Hiermit kann die Möglichkeit des selbstständigen Aufrufes neuer KI-Prognose für entsprechend qualifizierte Endnutzer freigegeben werden.

Das Verfahren wurde in der NiMo Endanwenderplattform beispielhaft für die Themen Messnetzoptimierung und Anomaliedetektion umgesetzt. In beiden Themenbereichen sind die entwickelten KI-Module regional übertragbar und können bei entsprechender Datenlage (Hintergrundinformationen sind in der Ontologie beschrieben) für alle im FROST®-Server integrierten Regionen eingesetzt werden.

Abb. 9 (links) zeigt die Umsetzung der Algorithmensteuerung für das Thema Messnetzoptimierung. In diesem Themenbereich stehen dem Anwender drei verschiedene KI-Module zur Verfügung: Vor der eigentlichen Prognose wird mit dem entsprechenden Datenbestand eine *Vorbereitung* durchgeführt. Auf Basis dieser Vorbereitung wird im *Ranking* die Wichtigkeit der Messstellen ermittelt. Mittels *Reduktion* können gezielt Teile eines Messnetzes eliminiert und die Fehlermetriken der verbleibenden Messstellen berechnet werden.

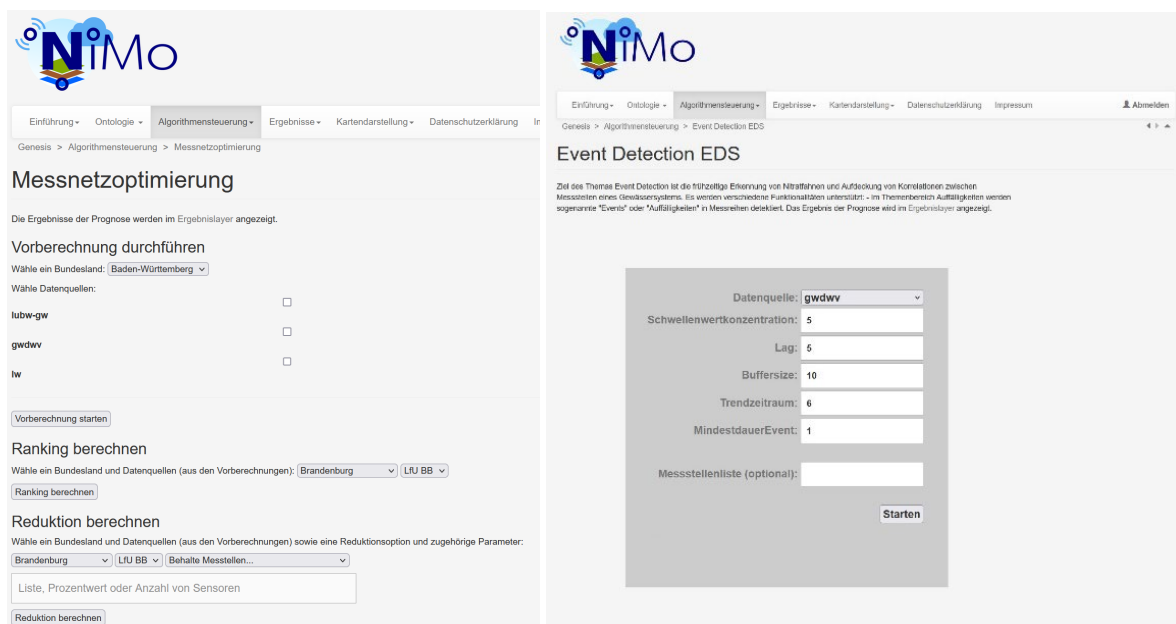


Abbildung 9: Beispiel für die Algorithmensteuerung

Abb 9 (rechts) zeigt die Algorithmensteuerung für das Event Detection System (EDS) des Themenbereichs Anomalieerkennung.

Die Ergebnisse der getriggerten Prognosen werden durch die KI-Module automatisch in die FROST® Instanz für die Verwaltung der Prognoseresultate integriert und stehen damit über die STA-Schnittstelle zur Abfrage und Visualisierung zur Verfügung.

Ergebnisdarstellung in WebGenesis

Prognose-Ergebnisse können auch für Endanwender ohne Anwendungskennnisse hinsichtlich der KI-Module freigegeben werden. In der Endanwenderplattform kann der Benutzer die Ergebnisse seiner eigenen oder früherer Berechnungen anzeigen. Die Ergebnisse sind im Menüpunkt Ergebnisse gesammelt.

Abb 10 zeigt Beispiele für verschiedene Möglichkeiten der Ergebnisdarstellung. Die Messstellenreduktion (Abb 10 links), markiert weniger wichtige Messstellen in rot im Gegensatz zu wichtigen in grün in einer Karte. Bei der allgemeinen Darstellung von Ergebnismessreihen kann der Nutzer die Ergebnisdaten nach Regionen filtern und in interaktiven Tabellen oder Diagrammen darstellen. Hierbei werden Optionen zum Zoomen und skalieren angeboten. Dies ermöglicht einem Fachexperten ein transparentes Verständnis der Daten und daraus abgeleitete Fachentscheidungen treffen zu können (Ergebnis des EDS Abb 10 rechts).

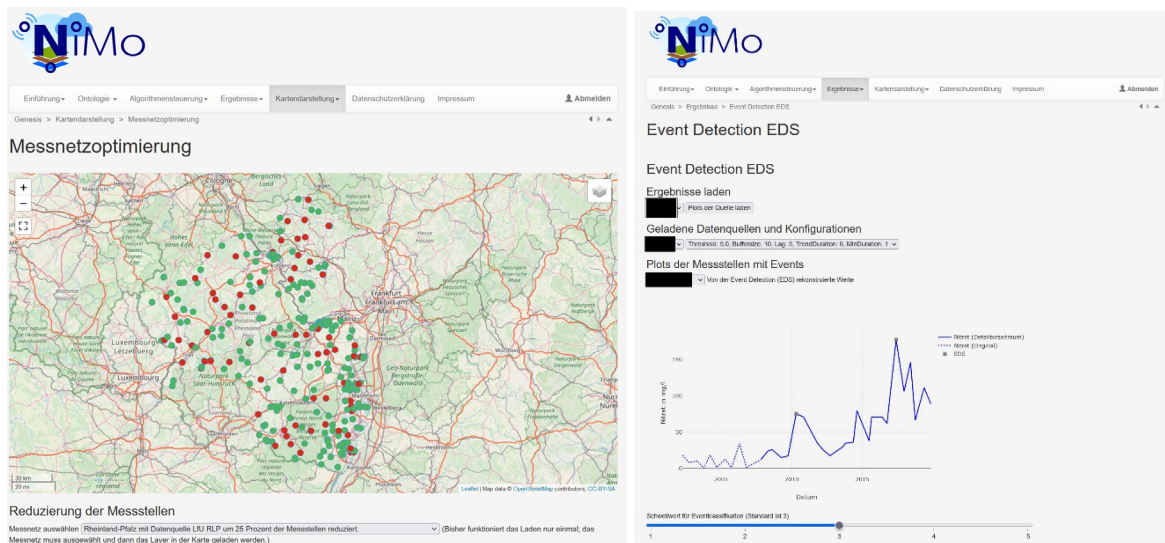


Abbildung 10: Beispiele für die Ergebnisdarstellung

Die Forschungsergebnisse des Themas Nitratregionalisierung beziehen sich nicht auf Zeitreihen, sondern auf eine regionale Berechnung von Nitratwerten, die in Karten visualisiert werden können. Die entsprechenden Prognoseergebnisse des Algorithmus werden demzufolge nicht in den FROST-Server® integriert, sondern über den GeoServer mittels dem OGC-Standards Web Map Server zur Verfügung gestellt. Die entsprechenden Layer können damit direkt in WebGenesis® Karten integriert werden.

Leuchtturmanwendungen für Pilotregionen

Die im Arbeitspaket „Anwenderspezifische Entscheidungsunterstützung“ erforschten generischen Methoden wurden benutzt um eine NiMo-Demonstrationsplattform auf Basis von WebGenesis® aufzubauen. Die im FROST-Server® integrierten Daten standen direkt den entwickelten KI-Methoden zur Verfügung. Nach erfolgreicher Integration einer KI-Methode kann diese aus technischer Sicht automatisch auf alle im Server verfügbaren Daten angewendet werden und ist somit auch direkt in neue Regionen verfügbar. Die Demonstrationsplattform funktioniert also über die ursprünglich angedachte Pilotregionen (LW; WAZ) hinaus für die Regionen für die weitere Bundesländer und Wasserversorger Daten geliefert haben.

Die integrierten Geobasisdaten für die Nitratregionalisierung wurden so ausgelegt, dass sie für eine deutschlandweite Auswertung geeignet wären. Limitationen sind hier bedingt durch die KI-Methode selbst, siehe Bericht des Partners AGW. Die Modelle der Messnetzoptimierung und der

Anomaliedetektion sind überwiegend datenbasiert und benötigen primär die Wasserqualitätsdaten aus dem FROST-Server®. Deswegen ist hier eine direkte Anwendung der Methoden in neuen Regionen möglich. Die Ergebnisse der Methoden sind allerdings stark abhängig von der jeweiligen Datenqualität und Datenverfügbarkeit in der Region. Entsprechende Hinweise für Endanwender wurden in die Beschreibung der KI-Methoden in der Ontologie aufgenommen.

Grundsätzlich können Informationen in der Demonstrationsplattform über ein Rechtekonzept freigegeben werden. Durch die erstmalige Integration auch nicht öffentlich verfügbarer Nitrat-Daten in einer gemeinsamen Plattform hat sich herausgestellt, dass ein entsprechendes Rechtekonzept auch für das Datenmanagement im FROST-Server® notwendig ist. Der FROST-Server® wurde zwischenzeitlich entsprechend erweitert. Die Funktionen sind open source verfügbar und können künftig zum Einsatz kommen.

Projektvernetzung, Übertragbarkeit und Nachhaltigkeit

Im Rahmen des Projektes wurden regelmäßig Kontakte zu Endanwendern gesucht. Zu Beginn waren vor allem die im Projektantrag vorgesehenen Anwendungspartner Zweckverband Landeswasserversorgung (LV), Wasser- und Abwasser-Zweckverband Niederrhein (WAZ) und Landesanstalt für Umwelt (LUBW) beteiligt.

Einerseits ist für die Entwicklung der KI-Modelle eine gute Datengrundlage essentiell und andererseits ist das Feedback von Endanwender für die Entwicklung von Demonstratoren sehr hilfreich. Deswegen wurden während der Projektlaufzeit mehrere Endanwenderworkshops offen für Interessierte veranstaltet. Hierbei konnten weitere Länder für die Datenbereitstellung gewonnen werden: Landesamt für Umwelt Brandenburg, Thüringer Landesamt für Umwelt, Bergbau und Naturschutz (TLUBN) mit Zugriff auf hochauflösende Datenlogger für Nitrat der SEBA Hydrometrie GmbH & Co. KG, der Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, das Landesamt für Umwelt Rheinland-Pfalz. Außerdem haben sich die Wasserversorger aus Baden-Württemberg bereit erklärt ihre Daten der Grundwasserdatenbank Wasserversorgung zur Verfügung zu stellen. Dadurch entstanden neue Kontakte zu Umweltverwaltungen und Firmen, die vor dem Projekt nicht bestanden haben.

Durch die neuen Daten konnte die Übertragbarkeit der NiMo Forschungsergebnisse auf neue Regionen getestet werden. Durch die Integration der Daten in das Datenmodell des FROST-Servers® standen die Daten direkt den KI-Modellen zur Verfügung. Die technische Übertragbarkeit der Architektur ist also gewährleistet. Die Übertragbarkeit der einzelnen KI-Methoden ist abhängig vom Anwendungsfall. Sie konnte für die Arbeitspakete Messnetzoptimierung und Anomaliedetektion ebenfalls nachgewiesen werden.

Wichtigste Position des zahlenmäßigen Nachweises

Abteilung ILT:

| Jahr | Personalkosten | Sachkosten | Geräte |
|------|----------------|------------|----------|
| 2020 | 19.084,33 | | |
| 2021 | 261.849,92 | 176,00 | |
| 2022 | 211.228,31 | 1.779,67 | 7.000,00 |
| 2023 | 145.925,50 | | 0 |

Abteilung UWR:

| Jahr | Personalkosten | Sachkosten | Geräte |
|------|----------------|------------|--------|
|------|----------------|------------|--------|

| | | | |
|------|------------|--------|--|
| 2020 | 10.078,47 | | |
| 2021 | 15.0294,90 | 174,08 | |
| 2022 | 175.901,43 | 179,88 | |
| 2023 | 61.308,28 | 198,00 | |

Notwendigkeit der geleisteten Projektarbeiten

Die Nutzung der für NiMo 4.0 erforderlichen Mittel war wie geplant notwendig für die Durchführung der erforderlichen Arbeiten welche zur Erreichung der Ziele des Projektes NiMo-4.0 umgesetzt werden mussten.

Im Bereich der Systemarchitektur, des Datenmanagements auf Basis offener Standards und anwenderspezifischen Entscheidungsunterstützung stellte sich die Notwendigkeit folgendermaßen da. Nur durch erstens den Aufbau des Datenmanagements auf Basis der SensorThings API, zweitens die Anpassung oder Weiterentwicklung existierender Komponenten oder die Erforschung neuer Komponenten für die Umsetzung des Datenmanagements, die sich auf Grund der Integration von KI-Methoden ergeben haben und drittens die Nutzung dieser Komponenten in der beispielhaften erstellten Endanwenderplattform zur Demonstration der Forschungsergebnisse in den Pilotgebieten konnten die Forschungsergebnisse der erforschten KI-Methoden nachgewiesen werden. Ein weiterer Anspruch für die technischen Komponenten war die Übertragbarkeit der definierten und getesteten Systemarchitektur auf neue Daten und Regionen um der Anforderung eines Leuchtturmprojektes zu genügen. Diese Übertragbarkeit konnte im Projektverlauf durch die problemlose Integration weiterer Daten neuer Endanwender und Nutzung der Methoden in neuen Regionen nachgewiesen werden.

Die Zuwendung von Ressourcen und Aufmerksamkeit auf die Erforschung und Anwendung von künstlicher Intelligenz in der Regionalisierung des Grundwasser Nitratgehalts und Eventdetektion ist gerechtfertigt, da diese Technologie das Potenzial hat, die bestehenden Probleme der konventionellen Methoden zu überwinden, wie die erreichten Ergebnisse zeigen. Künstliche Intelligenz ermöglicht eine schnellere Verarbeitung großer Datenmengen und kann komplexe Muster und Beziehungen erkennen, die von herkömmlichen Methoden möglicherweise übersehen werden. Darüber hinaus können die entwickelte künstliche Intelligenzmethoden auch die Generalisierungsfähigkeit verbessern und präzisere Vorhersagen für verschiedene Regionen ermöglichen. Daher ist die Erforschung und Anwendung von künstlicher Intelligenz in der Regionalisierung des Grundwasser Nitratgehalts ein vielversprechender Ansatz, um die Effizienz und Genauigkeit dieser wichtigen Umweltüberwachungsaufgabe zu verbessern. Viele herkömmliche Methoden zur Anomalieerkennung, wie Clustering oder Distanz-basierte Methoden, können hauptsächlich Ausreißer identifizieren. Durch KI-basierte Methoden können verschiedene Arten von Anomalien und Ereignissen in den Daten zur Grundwasserqualität identifiziert werden. Neben Punktanomalien (Ausreißer, die z.B. durch Defekte in den Sensoren entstehen) können die Modelle auch Kontextanomalien wie räumliche und zeitliche Ungleichmäßigkeiten sowie saisonale Veränderungen in den Parametern erkennen. Die potenziellen Vorteile rechtfertigen die Zuwendung von Ressourcen und Aufmerksamkeit auf die Entwicklung und Implementierung dieser Technologie in diesem Bereich.

Nutzen und Verwertbarkeit des Ergebnisses

Im Rahmen des Projektes konnten Verbesserungen an der open source Implementierung FROST® der SensorThings API vorgenommen werden, welche zukünftigen Nutzern der Implementierung

zur Verfügung stehen.⁵ Das Datenmodell der SensorThings API ist auf Zeitreihen fokussiert, eine Übertragung auf weitere Domänen und Anwendungsbereiche ist damit problemlos möglich.

Da die KI-Algorithmen in NiMo in Python entwickelt wurden, wurde im Speziellen ein Python FROST Client entwickelt und open source auf GitHub zur Verfügung gestellt.⁶ Mit Hilfe dieses Clients kann einfach in einer Python Implementierung auf Daten eines FROST-Servers zugegriffen werden.

Die Komponente PERMA[®] welche den 2. Teil der OGC SensorThings API „Part 2 – Tasking Core“ implementiert und mit dessen Hilfe die Ausführung der KI-Algorithmen gesteuert wurde, wurde für die Integration von KI-Algorithmen mit PyPerma weiterentwickelt und wird in nachfolgenden Forschungsprojekten wiederverwendet werden.

Insgesamt kann die aufgebaute NiMo-Architektur und die darin integrierten KI-Methoden als Blaupause für Projekte mit Ländern /Wasserversorgern dienen. Die entsprechenden Ergebnisse und Methoden liegen in einer Demonstrationsplattform vor. Eine Wiederverwendung der einzelnen Komponenten oder Forschungsergebnisse auch für interessante Teilaspekte ist möglich. Die Datenplattform wurde mit chemischen Qualitätsparametern vieler Länder aufgebaut. Prinzipiell ist die Nutzung der Plattform nicht auf Nitrat beschränkt, eine Übertragbarkeit auf andere Fragestellungen ist möglich. Prinzipiell ist die Integration weiterer KI-Methoden denkbar.

Da im Rahmen des Projektes einige Bundesländer und Wasserversorger auch nicht öffentliche Daten zur Verfügung gestellt und diese in die Plattform integriert und für die Entwicklung der KI-Methoden verwendet wurden, ist die Demonstrationsplattform nicht öffentlich verfügbar.

Die entwickelten Modelle für die Regionalisierung von Grundwasser Nitratgehalt und Ereignis/ Anomaliedetektion sind Werkzeuge zur Unterstützung von Umweltbehörden und -agenturen, Wasserversorger sowie Forschungseinrichtungen. Die können problemlos in anderen Bereichen eingesetzt werden, indem lediglich die entsprechenden Daten aus der jeweiligen Region genutzt werden. Zur Vereinfachung der Nutzung der Modelle wurde in Python Funktionen implementiert, die Daten mithilfe entwickelten Sensorthings API Python Client Bibliothek vom FROST[®]-Server importiert und die Daten in die Input Form der Deep Learning Modelle transformiert. So konnten die Deep Learning Modelle mit Trainingsdaten aus der Pilotregionen trainiert und anhand von entsprechenden Testdaten evaluiert werden

Fortschritt auf dem Gebiet des Vorhabens

Die SensorThings API (STA) ist ein offener Standard, der durch Verbesserungen im Gegensatz zum Sensor Observation Service (SOS), wie beispielsweise durch die Verwendung von JSON, überzeugt. Im Verlauf des Projekts konnten im Zusammenhang mit dem STA Standard neue Tools entwickelt werden, welche den Einsatz des Standards in der Praxis im Kontext der Verwendung von KI-Anwendungen ermöglicht.

Mit der Komponente PERMA[®] gab es bereits ein Tool, das sich um die automatische Ausführung von Algorithmen kümmert. Ähnliche Tools oder Ansätze die auf Verwendung eines öffentlichen Standards wie der SensorThings API basieren sind uns nicht bekannt. Von Interesse ist ggf. die Verfolgung der Weiterentwicklung des PAISE Modells, welches zur KI Integration in der Industriedomäne eingesetzt wird und die Analyse neuer Entwicklungen aus diesem Bereich⁷.

⁵ [FROST[®]-Server - Open-Source-Implementierung der OGC SensorThings API - Fraunhofer IOSB](#)

⁶ [Fraunhofer IOSB · GitHub](#)

⁷ [PAISE[®]: Das KI-Engineering Vorgehensmodell](#)

Da der Projektfokus auf KI-Algorithmen lag, galt es die Unterstützung für die Programmiersprache Python sicherzustellen, die im Feld der KI eine starke Bedeutung hat. Da PERMA ursprünglich in Java geschrieben wurde, stellt die neue Komponente PyPerma hierfür einen kompatiblen Adapter dar. Darüber hinaus wurde, nicht nur für den Tasking-Prozess und PERMA, sondern auch für das Abrufen und Schreiben von Daten über die SensorThingsAPI eine Python-Bibliothek benötigt. Mit dem Python Client wurde hier eine Open Source Lösung geschaffen, die auch über das Projekt hinaus gewartet und von Dritten verwendet werden kann.⁸⁹ Darüber hinaus wurden Möglichkeiten zum persistenten Ablegen von Modellen und Laden in einer Cloud-Plattform evaluiert.

Die konventionelle Methode der Regionalisierung von Grundwasser Nitratgehalt mit klassischen geostatistischen Methoden wie Kriging war in der Vergangenheit weit verbreitet. Diese Methoden sind jedoch sehr rechenintensiv und können Schwierigkeiten bei der Generalisierung der Ergebnisse aufweisen. Daher war es notwendig, die möglichen Einsatzmöglichkeiten von künstlicher Intelligenz in diesem Bereich zu untersuchen, um die Effizienz und Genauigkeit der Regionalisierung zu verbessern

Veröffentlichungen

Die folgenden Veröffentlichungen wurden im Rahmen von NiMo 4.0 mit Beteiligung des Fraunhofer IOSB erstellt:

Tanja Liesch, Julian Bruns, Andreas Abecker, Desirée Hilbring, Divas Karimanzira, Tobias Martin, Martin Wagner, Andreas Wunsch und Thilo Fischer (2020): Nitrat-Monitoring 4.0 – Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser. INFORMATIK 2020. Gesellschaft für Informatik, Bonn. (S. 1069-1079). DOI: 10.18420/inf2020_101.

Philipp Hertweck, Hylke van der Schaaf, Desiree Hilbring, Weis Jonas, Tanja Liesch und Matthias Budde (2021): Integration von KI-Algorithmen in Umweltinformationssysteme mittels SensorThings API. INFORMATIK 2021. Gesellschaft für Informatik, Bonn. DOI: 10.18420/informatik2021-024.

Katharina Emde, Matthias Budde, Thilo Fischer, Tobias Martin, Desiree Hilbring (2022): Interaktive Steuerung der Ausführung von KI-Algorithmen in Umweltinformationssystemen über OGC SensorThings. INFORMATIK 2022. Gesellschaft für Informatik, Bonn. DOI: 10.18420/inf2022_130.

Hilbring, Désirée; Becker, Kevin; Markus; Emde, Katharina; van der Schaaf, Hylke; Spandl, Horst; Tauber, Martina (2022): OData - Usage of a REST Based API Standard in Web based Environmental Information Systems. EnvirolInfo 2022. Bonn: Gesellschaft für Informatik e.V.. PISSN: 1617-5468. ISBN: 978-3-88579-722-7. pp. 53. Hamburg. 26.-30- September 2022

Hilbring, Desiree, Vogl Jonathan: Nutzung von Geodatenstandards für die Integration von KI-Algorithmen in IT-Systeme, Vernetzungstreffen der KI-Leuchttürme, Berlin, 23.03.2023

Karimanzira Divas, Weis Jonas, Wunsch Andreas, Ritzau Linda, Liesch Tanja, Ohmer Marc (2023), Application of machine learning and deep neural networks for spatial prediction of groundwater nitrate concentration to improve land use management practices Front. Water, 13 July 2023, Sec. Water and Artificial Intelligence, Volume 5 – 2023, <https://doi.org/10.3389/frwa.2023.1193142>, ISSN=2624-9375

⁸ [FROST®-Server - Open-Source-Implementierung der OGC SensorThings API - Fraunhofer IOSB](#)

⁹ [Fraunhofer IOSB · GitHub](#)

Divas Karimanzira, Linda Ritzau, Tobias Martin, and Thilo Fischer, "Advanced Spatio-Temporal Event Detection System for Groundwater Quality Based on Deep Learning." *Applied Ecology and Environmental Sciences*, vol. 11, no. 3 (2023): 79-90. doi: 10.12691/aees-11-3-2

Veröffentlichungen in Planung:

Matthias Budde, Desiree Hilbring, Jonathan Vogl, Daniel Dittmar, Andreas Abecker: NiMo 4.0: Enabling Advanced Data Analytics with AI for Environmental Governance in the Water Domain, *AI For Water Management*, 2024