

Geodatenanalyse 2: Machine Learning und Big Data

SS 2021

Tanja Liesch, Gabriel Rau, Andreas Wunsch, Jonas Weis



TOPs für heute

- ▶ Einordnung in den Studienplan
- ▶ Kleine Vorstellungsrunde
- ▶ Organisatorisches (Terminplanung und allgemeiner Ablauf, Infos zur Prüfungsleistung etc.)
- ▶ Testen der Technik
 - ▶ Einrichtung der Entwicklungsumgebung
 - ▶ Vorstellung Google Colab
- ▶ Beim nächsten Mal: Starten mit den Inhalten 😊

Profil „Hydro- und Ingenieurgeologie“

Stand: WS 2020-2021/SS 2021

1. SEMESTER	2. SEMESTER	3. SEMESTER	4. SEMESTER
Geowissenschaftliche Kernkompetenzen Pflicht 19 LP			MASTERARBEIT, 30 LP (Thema aus dem Fachbereich der Hydro- oder Ingenieurgeologie)
Numerische Methoden in den Geowissenschaften (*) 6 LP	Kartierkurs und Geodatenverarbeitung 8 LP	Projektstudie oder Berufspraktikum 5 LP	
Geowissenschaftliche Kernkompetenzen 36 LP (Pflicht für Profil Ing-Hydro)			
Hydrogeologie: Methoden und Anwendungen 7 LP		3D geologische Modellierung 5 LP	
Karst und Isotope 7 LP		Oberflächennahe Geothermie 5 LP	
Ingenieurgeologie: Massenbewegungen und Modellierung 5 LP			
Ingenieurgeologie: Gelände- und Labormethoden 7 LP			
Geowissenschaftliche Vertiefungen, 25 LP (5 aus 8 bzw. 9 Modulen)			
Aktuelle Forschungsthemen der Hydrogeologie und Ingenieurgeologie 5 LP			
Geodaten I: Programmierung und Geostatistik (*) 6 LP	Hydrogeologie: Gelände- und Labormethoden 5 LP	Hydrogeologie: Grundwasser-Modellierung 5 LP	
Felsmechanik und Tunnelbau (Import) 6 LP	Geodatenanalyse II – Big Data und Maschinelles Lernen 5 LP	Schadstoffe, Altlasten und Sanierungsmethoden 5 LP	
Geothermie: Energie- und Transportprozesse 5 LP	Geochemische Prozesse 5 LP		
Fachbezogene Ergänzungen, 10 LP (beispielhafte Kombination)			
Wahlmodul 5 LP		Wahlmodul 5 LP	
ca. 30 LP	ca. 30 LP	ca. 30 LP	30 LP
(*) Hinweise beachten!			
Summe 120 LP			

Profil „Hydro- und Ingenieurgeologie“

(voraussichtlich ab WS 2021-2022)

MASTERSTUDIENGANG ANGEWANDTE GEOWISSENSCHAFTEN / EXEMPLARISCHER STUDIENABLAUF

1. SEMESTER	2. SEMESTER	3. SEMESTER	4. SEMESTER
Geodatenanalyse I: Programmierung und Geostatistik 5 LP	Pflicht 20 LP	Projektstudie oder Berufspraktikum (unbenotet) 5 LP	MASTERARBEIT, 30 LP (Thema aus dem Fachbereich der Hydro- oder Ingenieurgeologie)
Ingenieurgeologie: Gelände- und Labormethoden 5 LP			
Hydrogeologie: Angewandte und Regionale Hydrogeologie 5 LP			
Wahlpflicht 55 LP (11 aus 14 Modulen)			
Ingenieurgeologie: Massenbewegungen und Modellierung 5 LP	Geodatenanalyse II: Big Data und Maschinelles Lernen 5 LP	Oberflächennahe Geothermie 5 LP	
Karsthydrogeologie & Exkursion 5 LP	Kartierkurs + GIS-Kartografie 5 LP	Hydrogeologie: Grundwasser-Modellierung 5 LP	
Felsmechanik und Tunnelbau (Import) 6 LP	Hydrogeologie: Gelände- und Labormethoden 5 LP	Schadstoffe, Altlasten und Sanierungsmethoden 5 LP	
Geothermie: Energie- und Transportprozesse* 5 LP	Hydraulik & Isotope 5 LP	3D geologische Modellierung 5 LP	
	Geochemische Prozesse 5 LP	Aktuelle Themen der Hydrogeologie und Ingenieur-geologie 5 LP	
Ergänzungen, 15 LP (beispielhafte Kombination)			
Wahlmodul 5 LP	Wahlmodul 5 LP	Wahlmodul 5 LP	
Summe 31 LP & 3 Prüfungen (bei Belegung der mit * gekennzeichneten Module)	Summe 29 LP & 6 Prüfungen (bei Belegung der mit * gekennzeichneten Module)	Summe 30 LP & 6 Prüfungen (bei Belegung der mit * gekennzeichneten Module)	30 LP

Summe 120 LP

Das sind wir ☺



Tanja Liesch

- ▶ Diplom-Geologin, Promotion in Hydrogeologie (2006)
- ▶ Arbeitsgruppenleiterin Numerische Modellierung und hydraulische Methoden
- ▶ Aktuelle Forschungsschwerpunkte:
 - ▶ Anwendung von KI auf hydrogeologische Fragestellungen
 - ▶ Statistische und numerische Methoden
- ▶ Aktuelle Projekte: NiMo 4.0, KARMA, SALAM2, BGR-KNN
- ▶ Lehrveranstaltungen: Hydraulische Methoden, Gelände- und Labormethoden, Grundwassermodellierung, Geodatenanalyse II
- ▶ Webseite: http://hydro.agw.kit.edu/21_158.php
- ▶ Google Scholar: <https://scholar.google.com/citations?user=0IzUw80AAAAJ&hl=de>
- ▶ Researchgate: https://www.researchgate.net/profile/Tanja_Liesch

Das sind wir ☺



Gabriel Rau

- ▶ Diplom-Ingenieur (2008), PhD (2012)
- ▶ Aktueller Forschungsschwerpunkt:
 - ▶ Quantifizierung von Untergrundeigenschaften durch die Antwort des Grundwassers auf Gezeitenkräfte
- ▶ Aktuelle Projekte: EC (SubTideTools), DFG (Tidal Subsurface Analysis)
- ▶ Lehrveranstaltungen in der Ingenieurgeologie: Wissenschaftliches Arbeiten und Kommunizieren, Gelände- und Labormethoden, Geodatenanalyse I + II
- ▶ Webseite: <https://hydrogeo.science>
- ▶ Google Scholar: https://scholar.google.com/citations?user=l_ehthwAAAAJ&hl=en
- ▶ Publons: <https://publons.com/researcher/489416/gabriel-c-rau/>

Das sind wir ☺



Andreas Wunsch

- ▶ B.Sc. + M.Sc. in AGW am KIT, Doktorand seit 2017
- ▶ Aktueller Forschungsschwerpunkt:
 - ▶ Zeitreihenvorhersagen mit KNN/DL (Grundwasserstände, Quellschüttungen etc.)
- ▶ Aktuelle Projekte: BGR-KNN, NiMo 4.0, KARMA
- ▶ Lehrveranstaltungen: Geodatenanalyse II
- ▶ Webseite: https://hydro.agw.kit.edu/21_172.php
- ▶ GitHub: <https://github.com/AndreasWunsch>
- ▶ Researchgate: <https://www.researchgate.net/profile/Andreas-Wunsch>

Das sind wir ☺



Jonas Weis

- ▶ 2017 B.Sc. in Geowissenschaften an der JGU Mainz
- ▶ 2020 M.Sc. in Angewandte Geowissenschaften an der RWTH Aachen
- ▶ Aktuelle Forschungsschwerpunkte:
 - ▶ Räumliche und zeitliche Vorhersage von Nitrat im Grundwasser mit Hilfe von maschinellem Lernen und künstlichen neuronalen Netzen
- ▶ Projekte: NiMo 4.0
- ▶ Webseite: https://hydro.agw.kit.edu/21_64.php

Wer seid Ihr?

- ▶ Studiengang BSc, am KIT oder anderer Uni?
- ▶ Profilbildung
- ▶ Motivation, das Modul zu belegen?
- ▶ Vorwissen im Bereich Big Data und Machine Learning?
- ▶ Erwartungen an die Veranstaltung? An uns?

- ▶ Das Modul **Geodatenanalyse II** wird dieses Jahr zum ersten Mal angeboten
- ▶ Inhalt ist der Umgang mit Big Data und der Einsatz von modernen Machine Learning Methoden in der Hydro- und Ingenieurgeologie – *vieles davon ist übertragbar auf andere Fragestellungen im geowissenschaftlichen Kontext!*
- ▶ Der Kurs besteht aus 12 inhaltlichen Terminen, jeweils mit Theorieteil, Übungen (Python) und Reflektion/Diskussion (plus Vorbesprechung plus 1 Termin Reserve/Zeit für Fragen)
- ▶ Grundkenntnisse in der Python-Programmierung werden vorausgesetzt → Modul Geodatenanalyse I
- ▶ Das Modul wird mit einer benoteten „Prüfungsleistung anderer Art“ abgeschlossen

Lernziele des Moduls

Am Ende des Moduls werden die Teilnehmer in der Lage sein:

- ▶ mit großen Datensätzen aus dem geowissenschaftlichen Bereich (z.B. Satellitendaten, Klimadaten) umzugehen.
- ▶ gängige Methoden der Zeitreihenanalyse auf hydro- und ingenieurgeologische Fragestellungen anzuwenden.
- ▶ Sie beherrschen grundlegende Verfahren des maschinellen Lernens.
- ▶ Sie können einfache Anwendungsfälle selbständig programmieren.

Was wird von Ihnen erwartet?

- ▶ Umfang: 5 LP = 150 h Arbeitsaufwand
- ▶ davon 42 h Präsenzzeit:
 - ▶ 14 Termine à 3 SWS im Semester = 42 h (davon 1 Termin Vorbesprechung und 1 Termin Reserve/Zeit für Fragen)
 - ▶ Präsenzzeit mit Anwesenheitspflicht!
- ▶ und 108 h Selbststudium:
 - ▶ Vor- und Nachbereitung der Vorlesung/Übungen (je ca. 1h für die 12 inhaltlichen Termine = 24 h)
 - ▶ Anfertigen der Prüfungsleistung anderer Art (der Rest, also 84 h = 10,5 Arbeitstage)

► Vorbereitung:

- An einigen Terminen wird der Theorieteil (kommentierte Vorlesungsfolien bzw. Screencasts) als Video zur Verfügung gestellt. Dies bitte jeweils als Vorbereitung auf die eigentliche Übung anschauen!

► Nachbereitung:

- Ggf. die Aufgaben aus der Übung zu Ende machen und/oder zur Übung bzw. Wiederholung eine weitere Aufgabe anfertigen.

Terminplan

	Datum	Dozent	Thema
Vorbesprechung	12.04.	Tanja & Gabriel	Vorbesprechung und Einrichtung der Entwicklungsumgebung
Termin 1	19.04.	Gabriel	Umgang mit und Analyse von Zeitreihen
Termin 2	26.04.	Tanja & Jonas	Grundlagen des maschinellen Lernens, EDA, Datenaufbereitung
Termin 3	03.05.	Gabriel	Frequenzanalyse von Zeitreihen
Termin 4	10.05.	Tanja & Jonas	Überwachtes Lernen (einfache ML Algorithmen), unüberwachtes Lernen (Clustering), Crossvalidation
Termin 5	17.05.	Gabriel	Erweiterte Kartografie mit Geopandas
Termin 6	31.05.	Tanja	Grundlagen neuronaler Netze, MLP, FFN
Termin 7	07.06.	Gabriel	Analyse von großen Datensätzen
Termin 8	14.06.	Tanja & Andreas	Rekurrente Neuronale Netze, Vorhersage von Zeitreihen mit CNN und LSTM
Termin 9	21.06.	Gabriel	Einführung in Google Earth Engine
Termin 10	28.06.	Tanja & Andreas	Hyperparameter-Tuning und Ensembles
Termin 11	05.07.	Gabriel	Anwendungen mit Google Earth Engine
Termin 12	12.07.	Tanja	Ausblick und interessante Anwendungen: XAI, Transfer-Learning, Autoencoder, SOM (kann ggf. auch wegfallen, wenn wir vorher nicht so weit kommen und einen weiteren Termin brauchen)
Reserve	19.07.	Tanja & Gabriel	Wiederholungen und Sprechstunde

Feedback

Während des Kurses:

- ▶ Fragen sind jederzeit willkommen!
- ▶ Wir bitten eventuelle Fehler im Kursmaterial zu melden
- ▶ Bei Problemen bitte direkt an die Dozenten wenden
- ▶ Lehrevaluation am Ende des Kurses
- ▶ Sprechstunde nach Absprache per E-Mail

- ▶ **Anwesenheit und aktive Teilnahme an den Übungen**
- ▶ **Prüfungsleistung „anderer Art“ in Form eines E-Portfolios:**
 - ▶ Bearbeitung mehrerer Aufgabenstellung (insgesamt 4) während und nach der Vorlesungszeit:
 - ▶ Teil 1 (Literaturrecherche und -besprechung): Öffnung nach 5 Terminen, Einreichungsfrist 4 Wochen.
 - ▶ Teil 2 (Datenrecherche und -besprechung): Öffnung nach 9 Terminen, Einreichungsfrist 2 Wochen.
 - ▶ Teil 3: Öffnung nach 11 Terminen, Einreichungsfrist 6 Wochen.
 - ▶ Teil 4 (Reflexion): Öffnung nach 13 Terminen, Einreichungsfrist 6 Wochen.
 - ▶ Darunter eine vorgegebenen Aufgabenstellung in *Python* (Teil 3)
 - ▶ Erstellen eines individuellen Workflows mit Code und Erklärung zur Analyse eines Geodatensatzes
 - ▶ Dokumentation in Form eines Jupyter Notebooks mit Visualisierung und Diskussion der Ergebnisse
 - ▶ Details zu den einzelnen Teilen bei Öffnung der Aufgabenstellung
 - ▶ Die Prüfungsleistung wird benotet und entspricht der Modulnote
 - ▶ **Prüfungsanmeldung in Campus: ab sofort, bis 30.06.2021!!!**

