



31. Workshop Umweltinformationssysteme (UIS 2024)  
"Digitalisierung für eine nachhaltige Planetare Zukunft"

# Prognose von Pegelständen mit Methoden des maschinellen Lernens und frei verfügbaren Daten

Grit Behrens, Daniel Marten, Levent Koch , HSBI Hochschule Bielefeld

Marcel Gaj, Landesamt für Natur, Umwelt und Verbraucherschutz NRW, Labor Minden

Landesamt für Natur,  
Umwelt und Verbraucherschutz  
Nordrhein-Westfalen



HS'BI  
Hochschule  
Bielefeld  
University of  
Applied Sciences  
and Arts

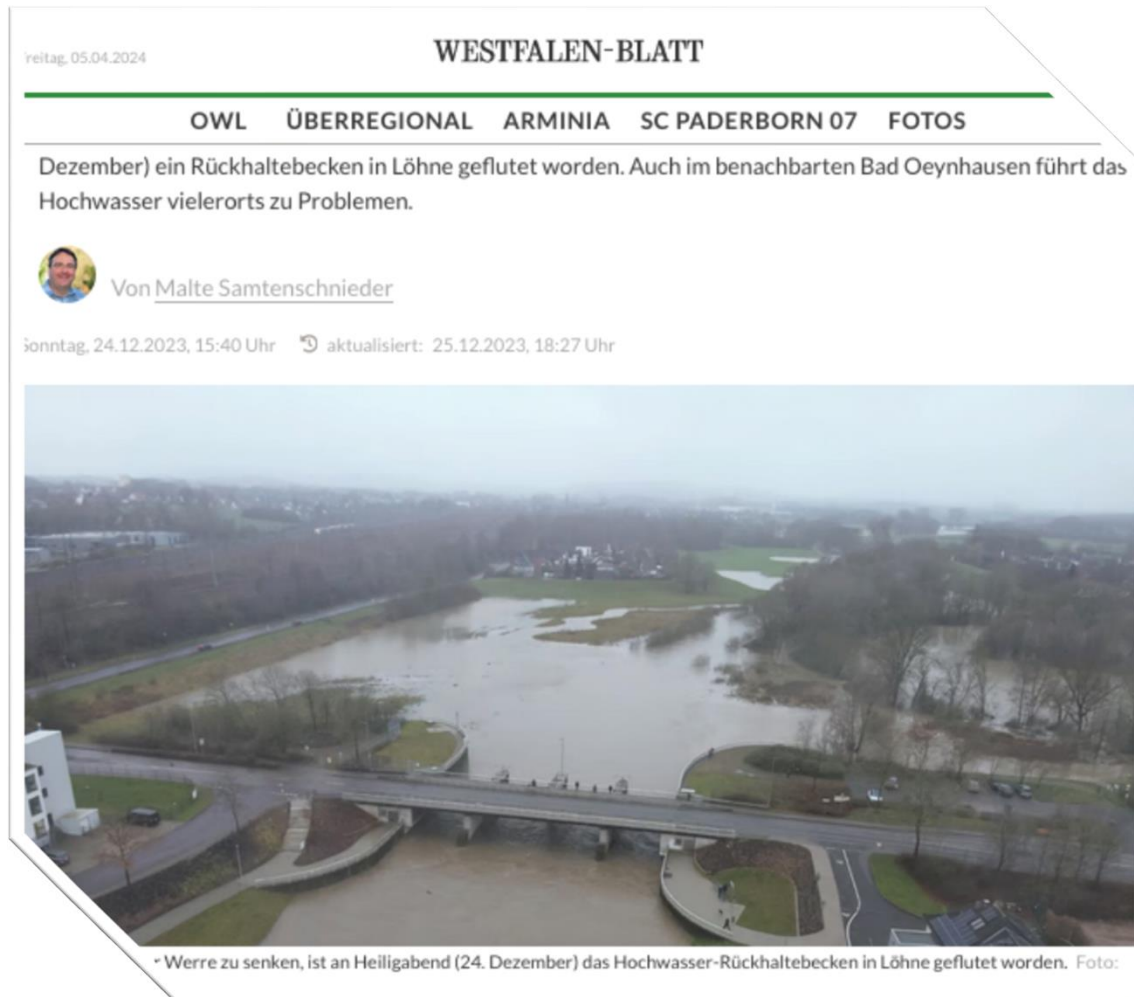
LANUV  
Kompetenz für ein  
lebenswertes Land

# Prognose von Pegelständen mit Methoden des maschinellen Lernens und frei verfügbaren Daten

- ▶ Einführung
- ▶ Zielsetzung und Forschungsfragen
- ▶ Vorgehensweise
- ▶ Datenquellen
- ▶ Daten-Vorverarbeitung und Datenanalyse
- ▶ Maschinelles Lernen
- ▶ Ergebnisse
- ▶ Diskussion und Fazit

# Einführung

- ▶ Rückhaltebecken werden zu klein im Werre-Gebiet
- ▶ 24.12.2023 - das Rückhaltebecken wird geflutet



Werre zu senken, ist an Heiligabend (24. Dezember) das Hochwasser-Rückhaltebecken in Löhne geflutet worden. Foto:

# Einführung

- ▶ Regelmäßige Hochwasser in der Werre-Region bereiten Gefahrenlage
- ▶ 16.12. 2023 Überschwemmungen in flussnahen Gebieten

-> Mitarbeiter des LANUV fragt Informatiker am Campus Minden nach Zusammenarbeit bei der Pegelstandprognose

WESTFALEN-BLATT

Freitag, 05.04.2024

OWL ÜBERREGIONAL ARMINIA SC PADERBORN 07 FOTOS

## Hoher Werre-Pegel bereitet Sorgen in Bad Oeynhausen

Bad Oeynhausen - Als „angespannt, aber vergleichsweise noch entspannt“, hat Stadtsprecher Volker Müller-Ulrich am zweiten Weihnachtstag (26. Dezember) die Hochwasserlage in Bad Oeynhausen bezeichnet. Besonders der hohe Werre-Pegel sei besorgniserregend.

Von  Malte Samtenschnieder

Dienstag, 26.12.2023, 17:16 Uhr  aktualisiert: 26.12.2023, 20:25 Uhr



... ist die Weser über die Ufer getreten. Der Amanda-Anleger (Foto), die Sportplätze von Rot-Weiß Rehme und die ...e Weserhütte sind unter anderem betroffen. Foto: Malte Samtenschnieder

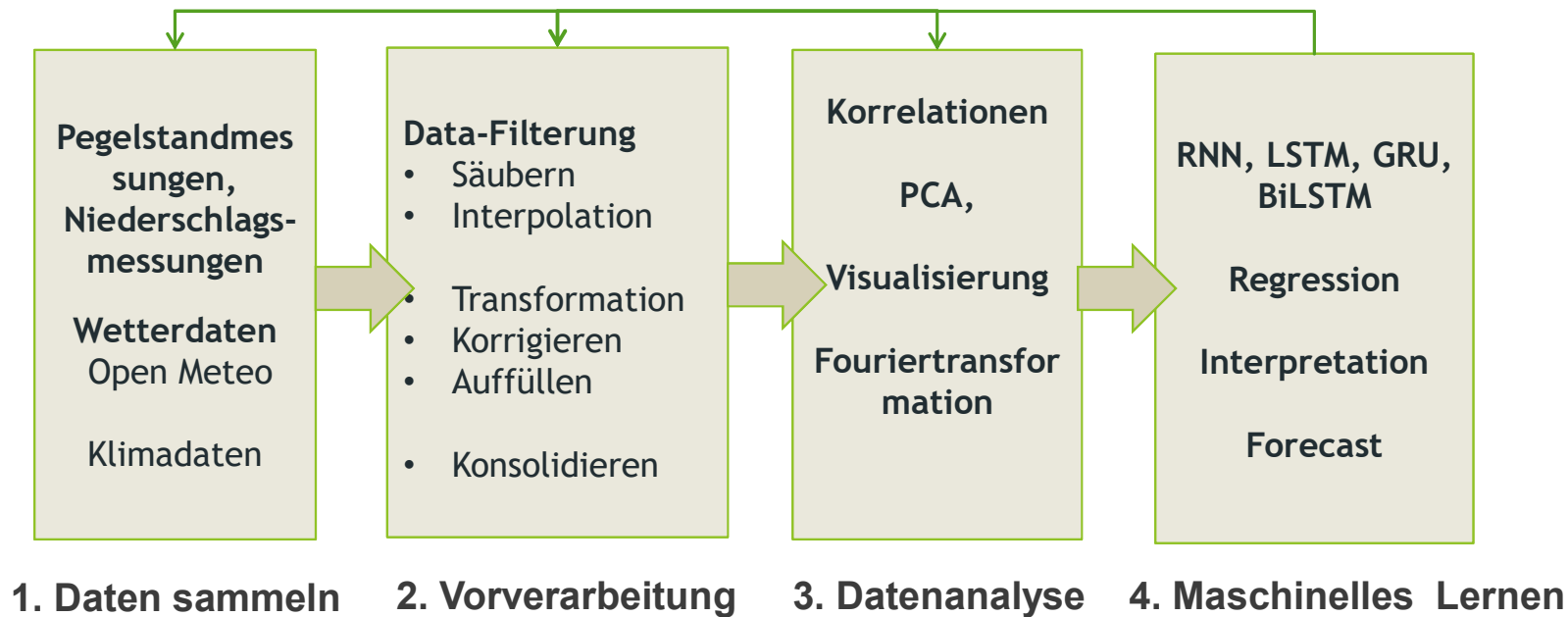
# Zielsetzung und Forschungsfragen

- ▶ Es soll ein System entwickelt werden, welches präzise Vorhersagen der Wasserpegelstände anhand von frei verfügbaren Daten liefert.
- ▶ Verfügbar sind Pegelstandmessungen und Klimadaten des LANUV sowie open-source Daten im Internet

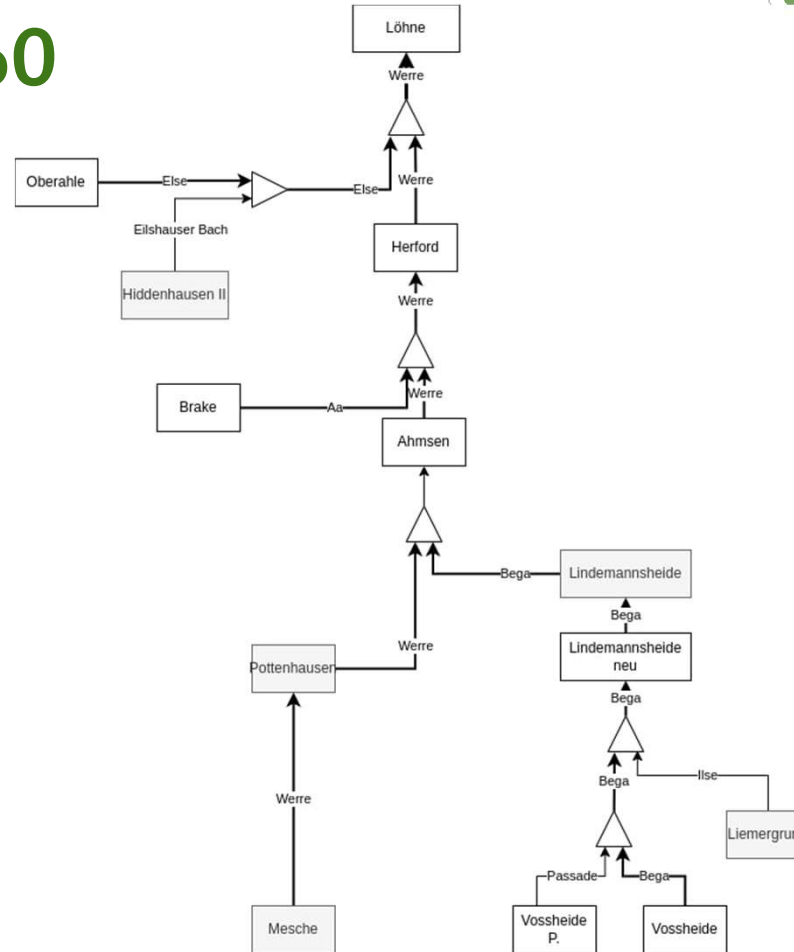
## Forschungsfragen:

1. Welche Merkmale/Daten sind für die Vorhersage der Pegelstände relevant?
2. Wie groß müssen die Zeiträume sein, um das Zuflussmodell mit Pegelstandsmessungen nutzen zu können?
3. Können die Pegelstände rein datengetrieben vorhergesagt werden? Welcher ML-Algorithmus eignet sich am besten?

# Vorgehensweise



# Datenquellen- Messstationen des LANUV für Pegelstände ab 1960

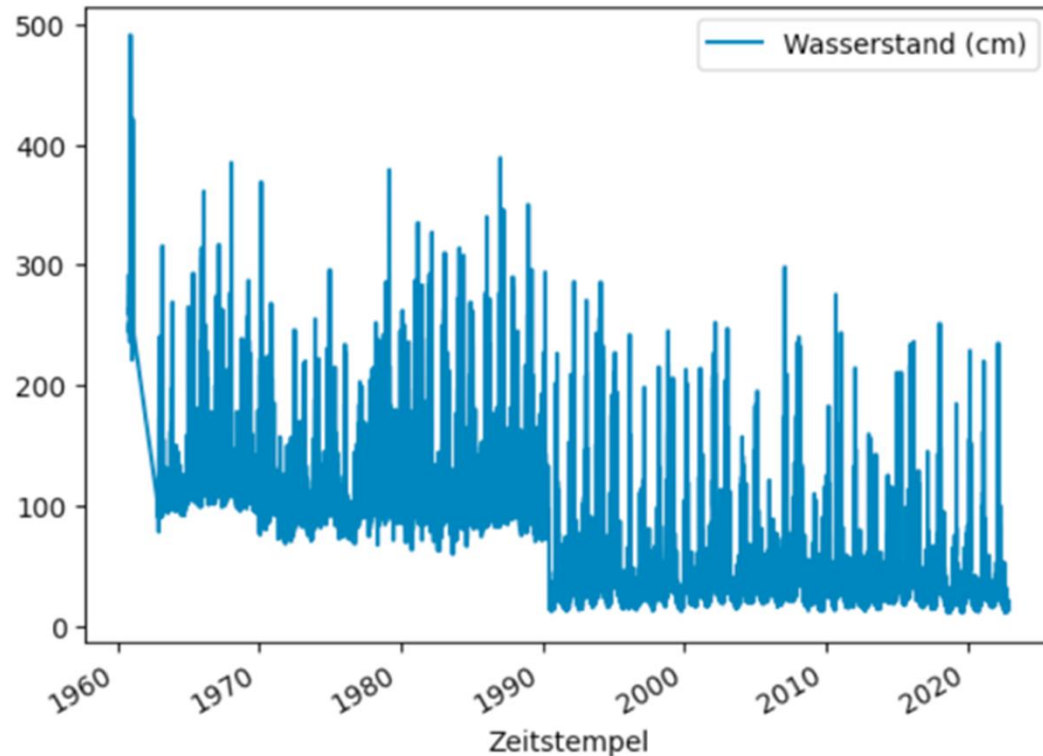


# Datenquellen: Messstationen des LANUV

- ▶ Es gibt Datenlücken!
- ▶ Beispielhafte Implementierung für einige Messstationen mit Daten für alle Zuflüsse im gleichen Zeitraum
- ▶ Auswahl ab 1996

## LANUV:

- ▶ Beispiel Daten für Pegelstandsmessungen ab 1960 in Messstation Oberahle
- ▶ Zusätzlich gibt es Niederschlagsdaten an den Messstationen

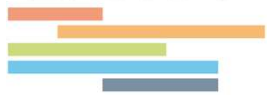




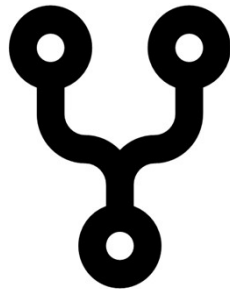
# Datenquellen: Openmeteo

## Daten Konsolidierung über GeoCode:

**LANUV**



Niederschlag,  
Pegelstände  
Zeitstempel



CSV-Datafiles



## OpenMeteo API - stündliche Wetterdaten ab 1996 (Auszug):

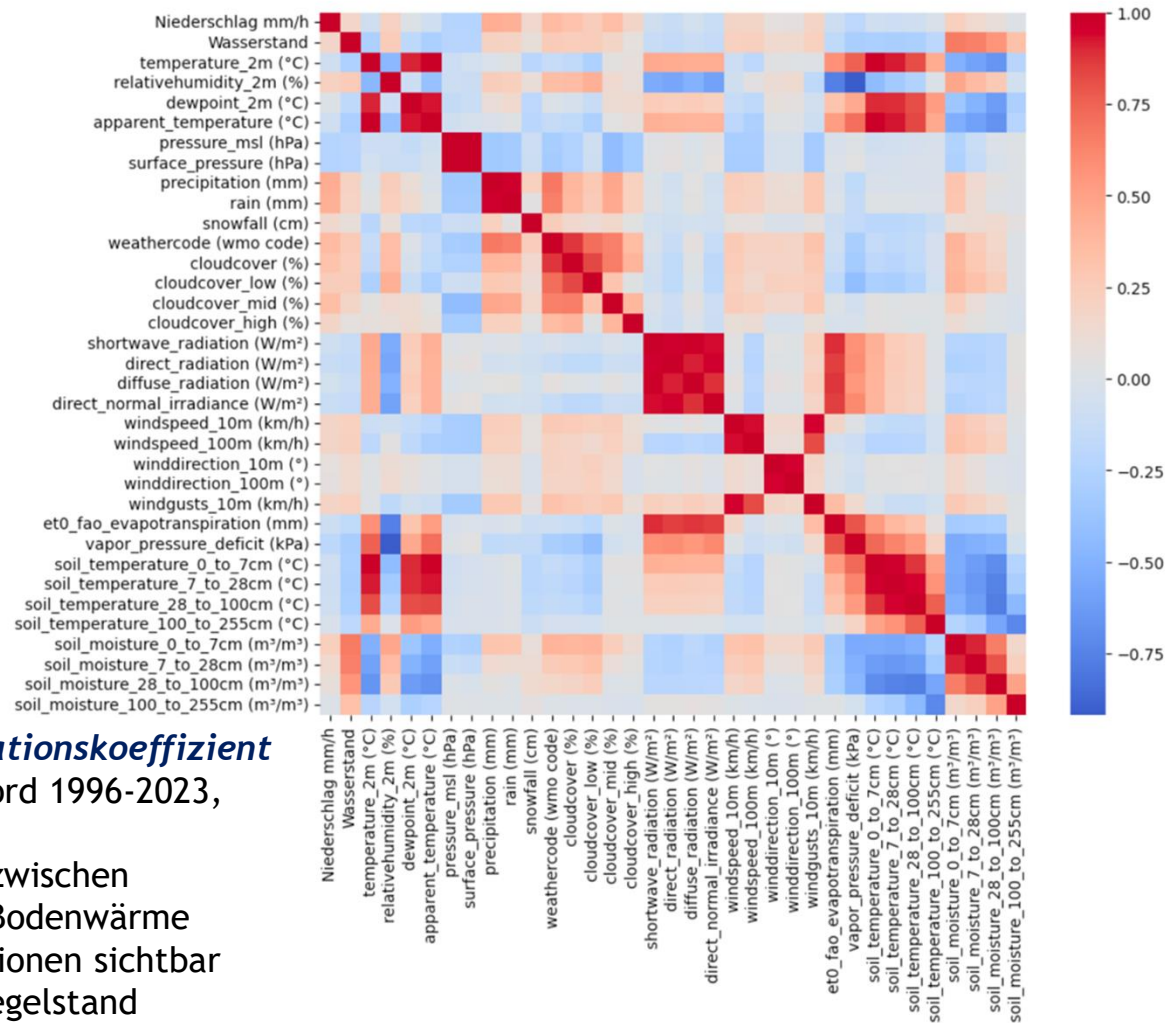
- Zeitstempel
- Temperatur (2m Höhe),
- relative Luftfeuchtigkeit (2m Höhe),
- Taupunkt (2m Höhe),
- Gefühlte Temperatur,
- Luftdruck auf Meereshöhe,
- Niederschlag, Regen, Schneefall,
- Bewölkung
- Solare Einstrahlung
- Windgeschwindigkeit, Windrichtung
- Bodentemperatur (0 bis 7 cm Tiefe),  
Bodentemperatur (7 bis 28 cm Tiefe),  
Bodentemperatur (28 bis 100 cm Tiefe) ,  
Bodentemperatur (100 bis 255 cm Tiefe),
- Bodenfeuchtigkeit (0 bis 7 cm Tiefe),  
Bodenfeuchtigkeit (7 bis 28 cm Tiefe) ,  
Bodenfeuchtigkeit (28 bis 100 cm Tiefe) ,  
Bodenfeuchtigkeit (100 bis 255 cm Tiefe)

# Datenvorberarbeitung und -analyse

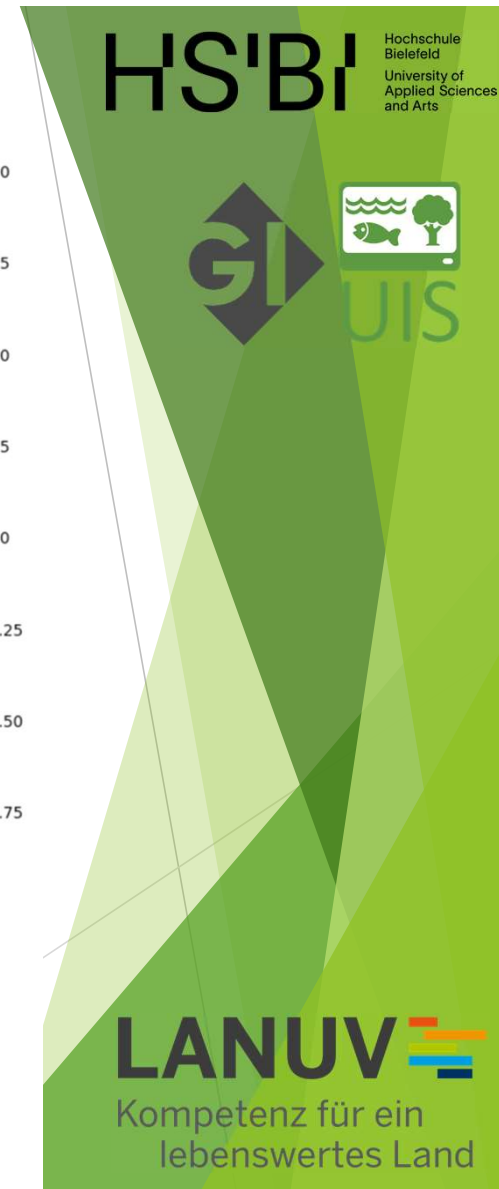
## Vorgehen:

- ▶ Ziel **Reduktion der Daten auf relevante Features** mit dem Analyseziel Hochwasserereignis
- ▶ Untersuchung der **Korrelationen der Features** alle untereinander und mit der Zielvariablen Pegelstand
- ▶ **Heatmaps** über alle Daten von OpenMeteo und LANUV gesamter Zeitraum **1996-2023**
- ▶ **Heatmaps** für Hochwasserereignisse **1.01.2019 - 1.01. 2020**

# Datenanalyse : HeatMap für alle Zeitpunkte

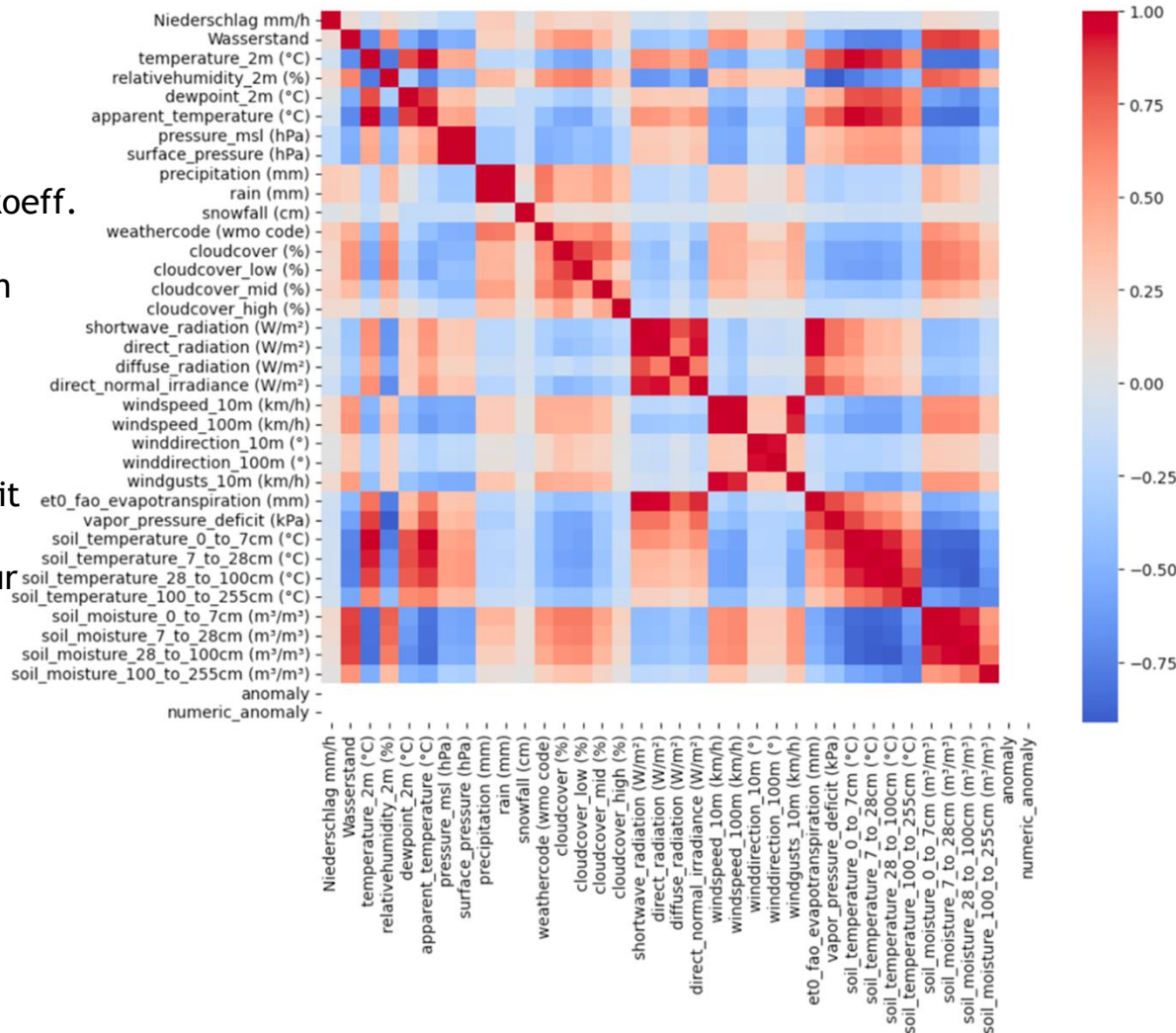


- **Spearman's Korrelationskoeffizient** für Messstelle Herford 1996-2023, **alle Features**
- z.B. Korrelationen zwischen Bodenfeuchte und Bodenwärme
- Aber kaum Korrelationen sichtbar zw. Features und Pegelstand



# Datenanalyse : HeatMap für Hochwasserereignisse (1.01.2019- 1.01.2020)

- Spearmans Korrelationskoeff. Messstelle Herford alle Features - nur Anomalien
- Korrelation zwischen Wasserpegel mit
  - Temperatur 2m
  - Rel. Luftfeuchtigkeit
  - Taupunkt
  - gefühlte Temperatur
  - Wolkendecke
  - Windgeschw.
  - Windstöße
  - Bodentemperatur
  - Bodenfeuchte

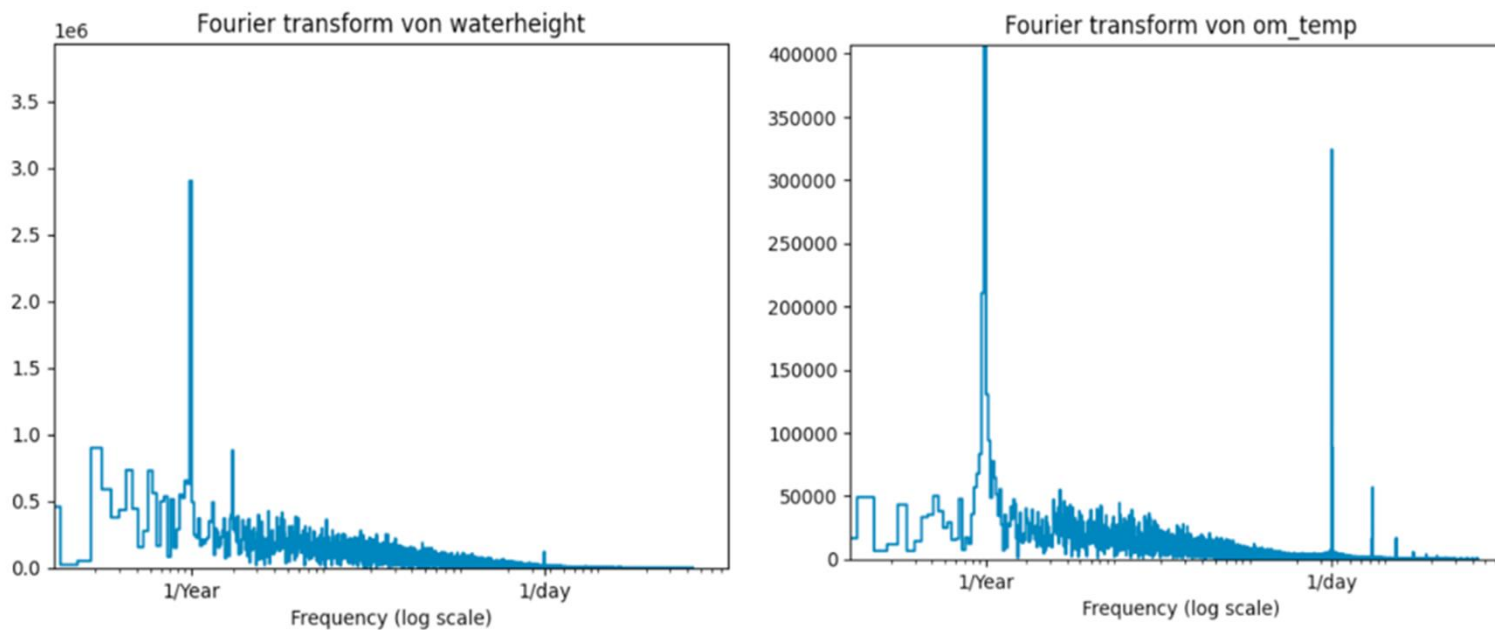


# Datenanalyse : weitere Datenreduktion

- Korrelationsanalyse
- Grafische Auswertung der Pegel im Flussverlauf (Hohe **Pegelstände** bei Hochwasser brauchen von erster Messstelle in Vossheide bis zur letzten Messstelle in Löhne maximal-> **Lag 30 Stunden**)
- Grafische Auswertung für **Lag** starker **Niederschlag** bis erhöhter Wasserstand dauert bis zu **6 Stunden**
- Konsultation des Experten (Weglassen von Features: *gefühlte Temp.*, *Cloudcover*, *Bodentemp ab 1 m*, *Bodenfeuchtigkeit ab 1 Meter*)
- Analyse der Zeitstempel - **Konsolidierung** auf **stündliche Werte** von OpenMeteo (Messen der Wasserstände und Niederschläge 15 minütig)

# Datenanalyse : Reengineering des Zeitstempels

- **Fouriertransformationen der Features** ergaben ein zyklisches Verhalten im Jahresverlauf und innerhalb von 24 Stunden (**Tag** und **Jahr**)
- Ersetzen des Zeitstempels mit **Sinus-und Cosinuswerten von Jahr und Tag**

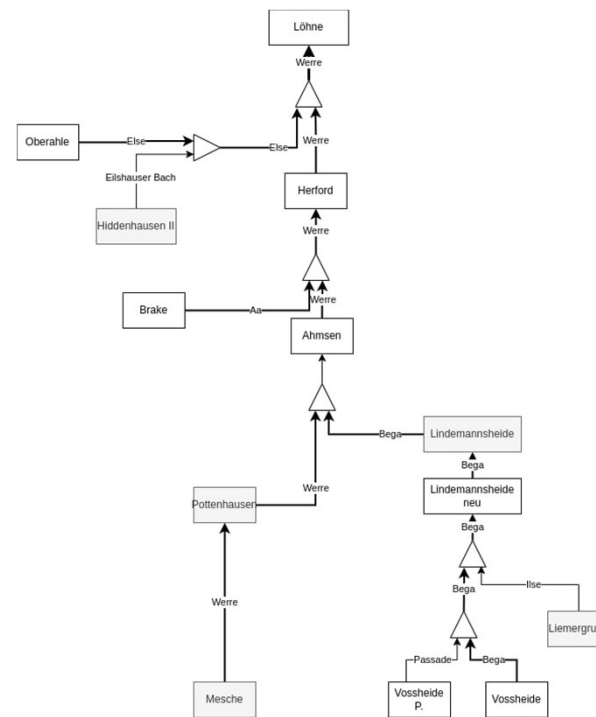


Beispiel der Fouriertransformation für alle Daten 1996-2023 von Wasserstand und Temperatur

# Maschinelles Lernen - Featurevektor

## Featurevektor für ML-Experimente stündliche Werte:

- ▶ Time:
  - ▶ day sin/cos
  - ▶ year sin/cos
- ▶ Pegelstände mit Lag 30
- ▶ Niederschläge mit Lag 6 (historische Daten fuzzifiziert mit 5% Fehler)
- ▶ Mittlere Bodentemperatur bis 1 m
- ▶ Mittelere Bodenfeuchtigkeit bis 1 m
- ▶ Temperatur
- ▶ Luftdruck
- ▶ Relative Luftfeuchtigkeit



# Maschinelles Lernen - Experimente

- ▶ **Prognose der Pegel 24 Stunden im Voraus für jeweils eine Messstation** anhand der 30 stündigen Lags der Wasserstandsmessungen aus den jeweiligen Zuflüssen und den 6 stündigen Lags der Open-Meteo-Daten
- ▶ Durchführen von Experimenten mit Rekurrenten Neuronalen Netzen davon **GRU** und **LSTM**
- ▶ **Optimierungsexperimente** über verschiedenen Architekturen und **Hyperparameter**
- ▶ Beste Prognoseergebnisse mit rel. einfachem **dreischichtigem LSTM**

|            |  |         |                     |
|------------|--|---------|---------------------|
| lstm_input |  | input:  | [(164621, 24, 206)] |
| InputLayer |  | output: | [(164621, 24, 206)] |

|      |      |         |                   |
|------|------|---------|-------------------|
| lstm |      | input:  | (164621, 24, 206) |
| LSTM | tanh | output: | (164621, 32)      |

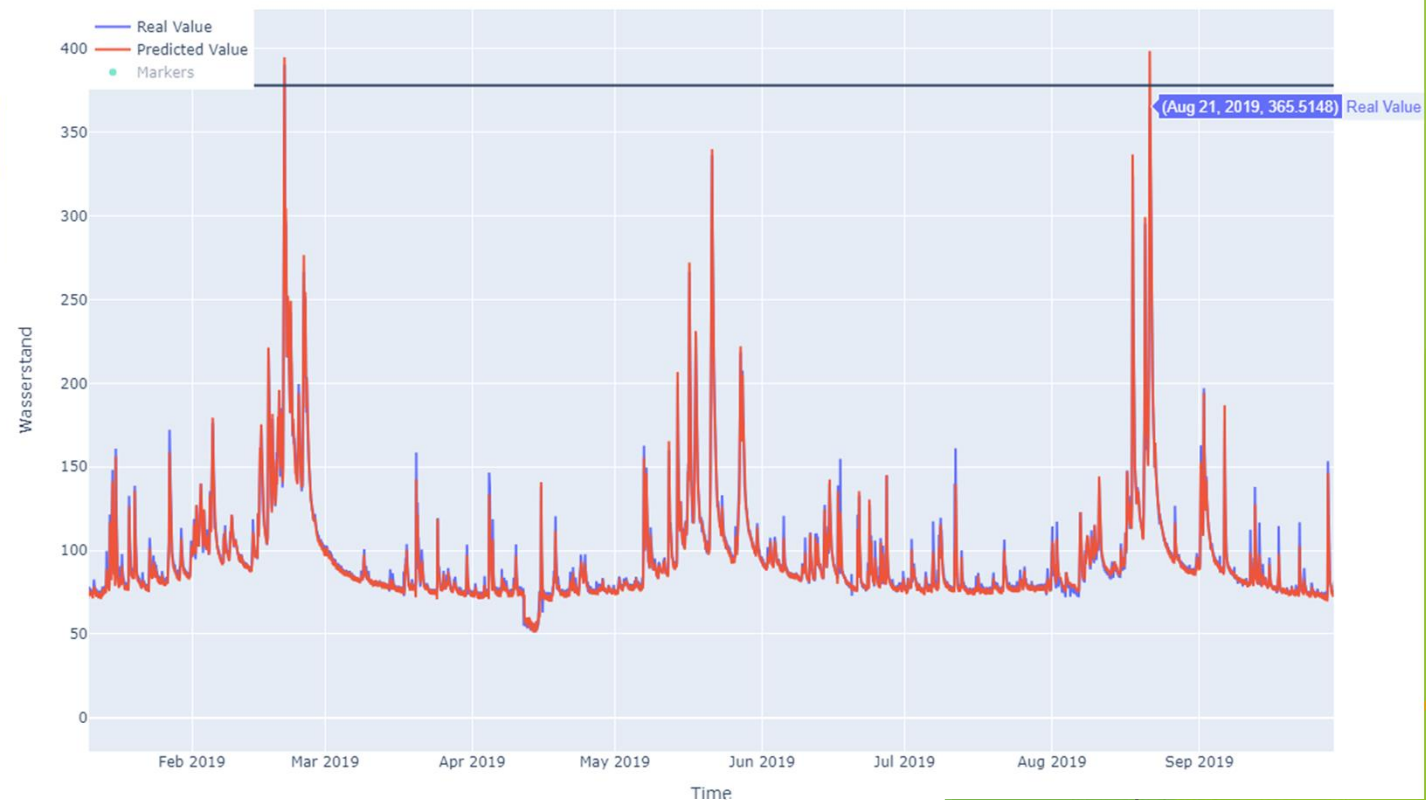
|       |        |         |              |
|-------|--------|---------|--------------|
| dense |        | input:  | (164621, 32) |
| Dense | linear | output: | (164621, 24) |



# Maschinelles Lernen - Ergebnisse für eine Messstation- Löhne

|                       | MSE   | RMSE | MAE  |
|-----------------------|-------|------|------|
| Testdatensatz         | 9.75  | 3.12 | 1.98 |
| Validierungsdatensatz | 13.17 | 3.63 | 2.52 |

Wasserstand Prediction



# Diskussion und Fazit

## Forschungsfragen:

### **1. Welche Merkmale/Daten sind für die Vorhersage der Pegelstände relevant?**

- Vorhergesagter Niederschlag mit Lag 6
- zyklisches tägliches und jährliches Verhalten der Einflussfaktoren
- Pegelstände der Messstelle und der Zuflüsse mit Lag 30

### **2. Wie groß müssen die Zeiträume sein, um das Zuflussmodell mit Pegelstandsmessungen nutzen zu können?**

- grafische Auswertung der Pegelstände von Zuflüssen und Messstelle selbst
- 30 Stunden für Messstellen an der Werre

### **3. Können die Pegelstände rein datengetrieben vorhergesagt werden? Welcher ML-Algorithmus eignet sich am besten?**

- mit hoher Präzision möglich
- große Abhängigkeit von Güte, Messsdichte und Durchgängigkeit der Messdaten
- Größere Lags der Dateneingaben in einfachen LSTM's mit z.B. drei Layern sind am besten geeignet.

# Vielen Dank für Ihre Aufmerksamkeit!



**HS'BI**  
Hochschule  
Bielefeld  
University of  
Applied Sciences  
and Arts

