

Wat gaat AI opleveren voor het NHI?

NHI-dag 18 januari 2024 -

Blik op de Toekomst met het oog op Innovatie, Verbinding & Community Building

Bennie Minnema

Met dank aan

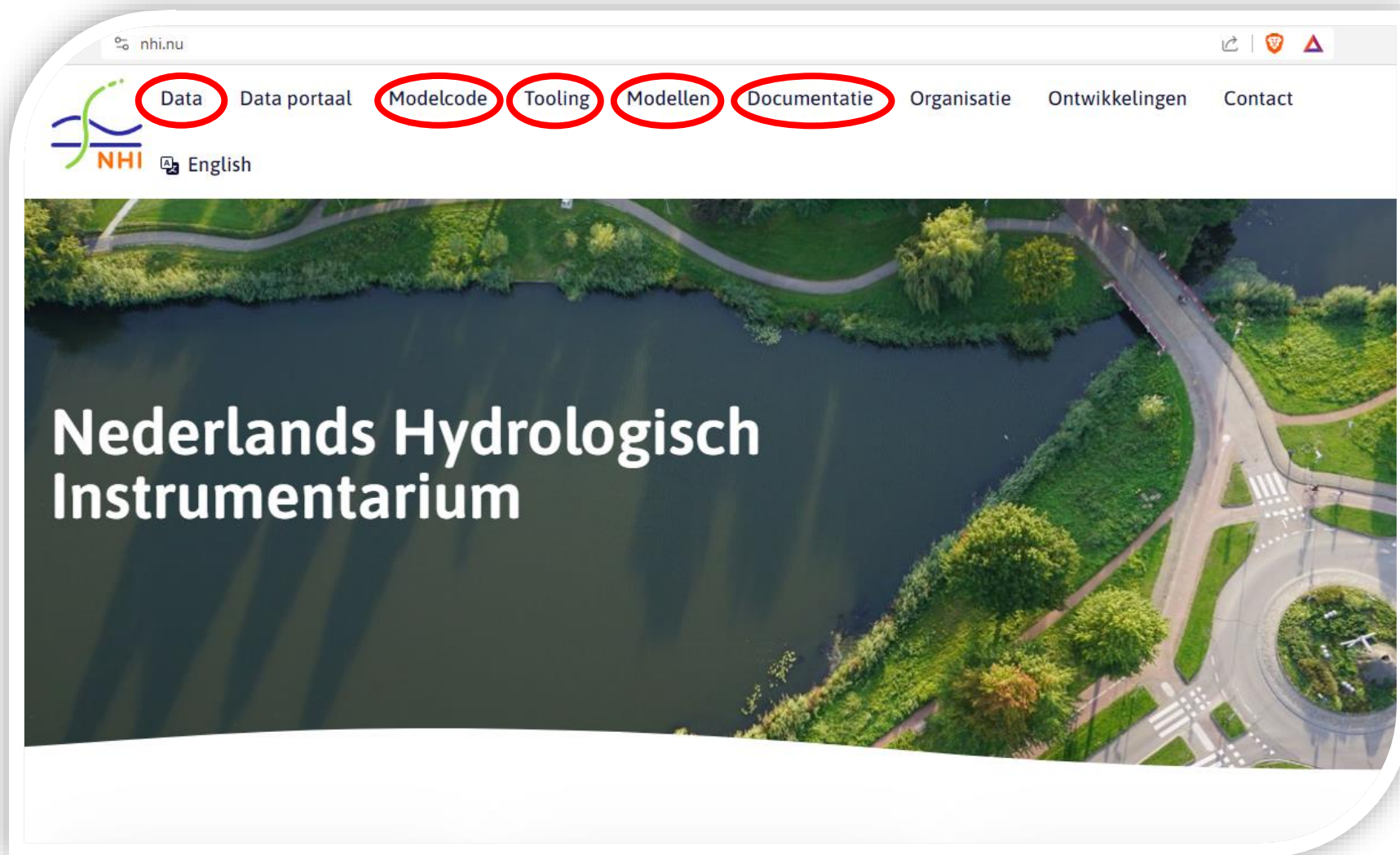
Antonio Moreno Rodenas – Enabling Technologies Data Science

AI: Artificiële of kunstmatige Intelligentie

Definities (bron: Chat-GPT):

- **Artificiële Intelligentie (AI)** is de theorie en ontwikkeling van computersystemen die taken kunnen uitvoeren waarvoor normaal *menselijke intelligentie* nodig is, zoals visuele waarneming, spraakherkenning, besluitvorming en vertaling tussen talen. ... en zichzelf herhaaldelijk kan verbeteren op basis van vergaarde informatie.
- **Machine learning:** een subset van AI waarbij algoritmen in de loop van de tijd hun prestaties kunnen verbeteren als ze aan meer gegevens worden blootgesteld.
- **Deep learning:** een subset van machine learning waarbij meerlaagse neurale netwerken leren van enorme hoeveelheden gegevens
- **Neuraal netwerk:** een reeks algoritmen die worden gebruikt als proces bij machinaal leren en die patronen en relaties in grote hoeveelheden gegevens kunnen herkennen. Neurale netwerken gebruiken een logische structuur geïnspireerd door het menselijk brein en vormen de basis voor deep learning-algoritmen.

Scope van deze presentatie: AI $\cap$ NHI



Artificiële Intelligentie als onderdeel van 'data science'

AI-'werkgebieden':

1. Generatieve AI
 - Large Language Models (LLM's)
 - b.v. ChatGPT van OpenAI: geavanceerde chatbot
2. Het AI-gebaseerd verwerken en analyseren van grote (meet)datasets
 - Foto's, video's, remote sensing data
3. 'Data-driven' modelbouw
 - Classificatie
 - Regressie
 - Clustering
4. Emulators of 'Surrogate Models'
 - Nabootsen van bestaande numerieke modellen

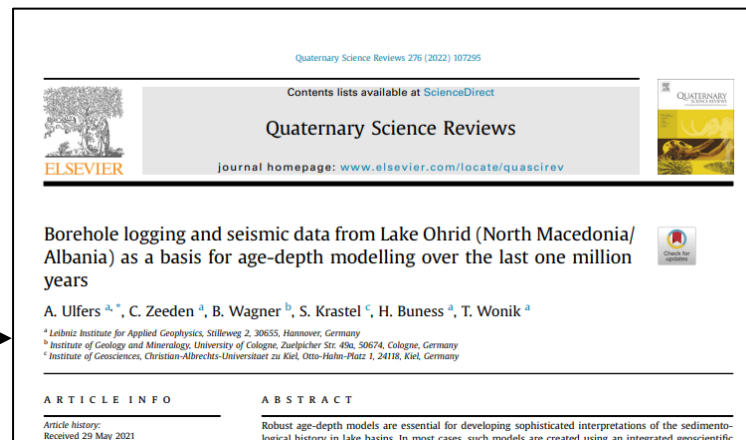
Artificiële Intelligentie als onderdeel van 'data science'

1. Generatieve AI
 - Large Language Models (LLM's)
 - b.v. ChatGPT van OpenAI: geavanceerde chatbot
2. AI-gebaseerd verwerken en analyseren van grote (meet)datasets
 - Foto's, video's, remote sensing data
3. 'Data-gedreven' modelbouw
 - Classificatie
 - Regressie
 - Clustering
4. Emulators of 'Surrogate Models'
 - Nabootsen van bestaande numerieke modellen

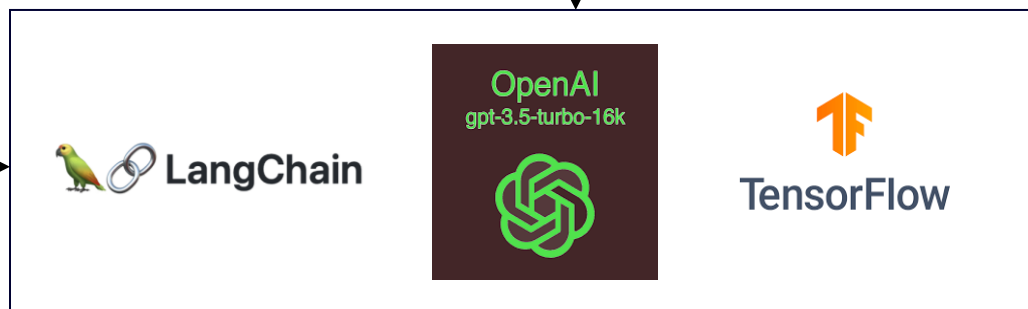
Extractie van grondwater data uit 'grey literature'

Data mining based on OpenAI API (ChatGPT)

Document with text/images
(e.g. reports, engineering files, scientific articles)



Python script
with APIs



Contextual query:
(e.g. give me all borehole and seismic essays in Albania)

Catalogued and
georeferenced data

Deltares



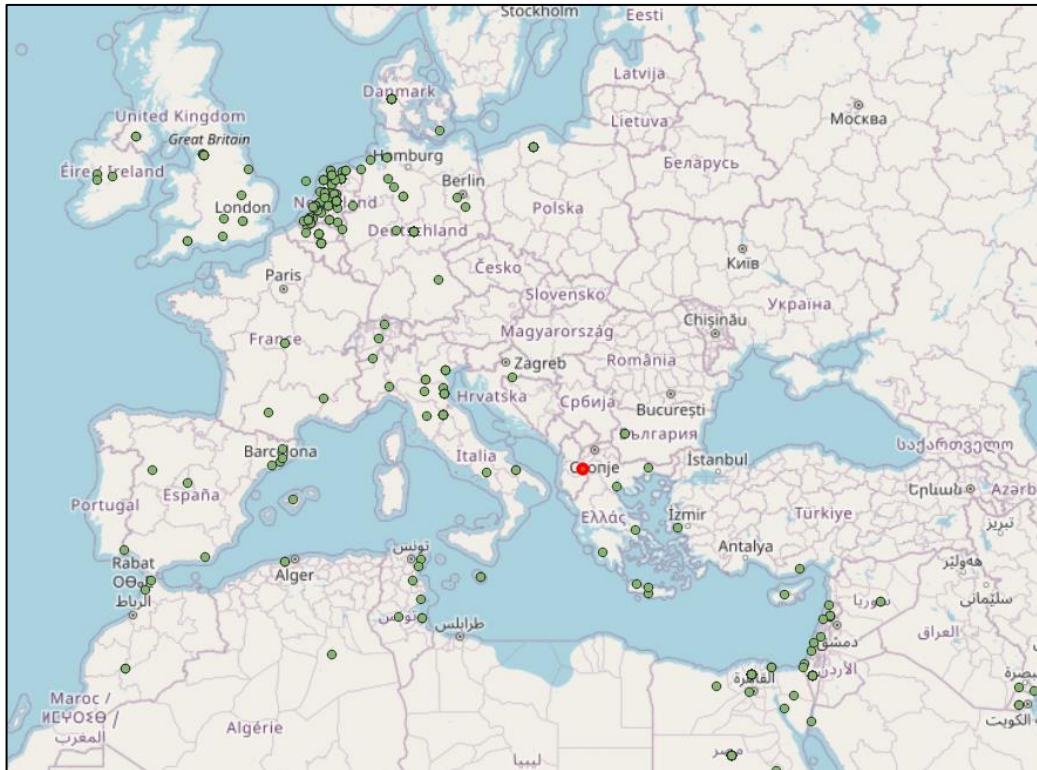
- Custom data mining routine developed for extraction of groundwater model data
- Based on **large language models** and trained **neural network image classifier**
- Automated extraction of parameters of interest, including georeferencing
- Automated classification of image type (work in progress!)
- Computation time ~6hrs per 1000 documents
- API calls and prompts handled by LangChain
- Tested on 2000+ documents

Data_Mine_V1	
title	Borehole logging and seismic data from Lake Ohrid (North...
(Derived)	
(Actions)	
field_1	1
title	Borehole logging and seismic data from Lake Ohrid (North...
year	2022
country	North Macedonia/Albania
region	Lake Ohrid
url	https://doi.org/10.1016/j.quascirev.2021.107295
keywords	['age-depth modelling', 'downhole methods', 'seismic inte...
data_links	[]
Y_country	41,139945980000000
X_country	20,065076829999999
Y_region	41,060115809999999
X_region	20,731275560000000
directory	C:\Users\king_je\Projects\Text_Mining\pdfs\1-s2.0-S02773...

Extractie van grondwater data uit 'grey literature'

Output in GIS

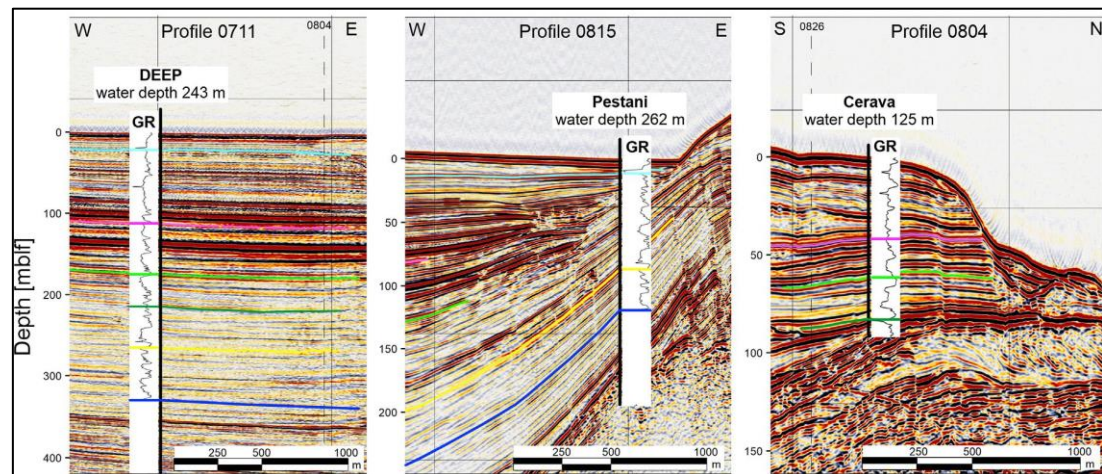
1. Select georeferenced data



2. Check information

▼ Data_Mine_V1	
▼ title	Borehole logging and seismic data from Lake Ohrid (North...
▶ (Derived)	
▶ (Actions)	
field_1	1
title	Borehole logging and seismic data from Lake Ohrid (North...
year	2022
country	North Macedonia/Albania
region	Lake Ohrid
url	https://doi.org/10.1016/j.quascirev.2021.107295
keywords	['age-depth modelling', 'downhole methods', 'seismic inte...
data_links	[]
Y_country	41,139945980000000
X_country	20,065076829999999
Y_region	41,060115809999999
X_region	20,731275560000000
directory	C:\Users\king_je\Projects\Text_Mining\pdfs\1-s2.0-S02773...

3. Extract automatically classified images



Sneller betere softwarecode schrijven

Gemakkelijker schrijven van (tooling)software:

- [GitHub Copilot](#): ontwikkeld door GitHub in samenwerking met OpenAI, werkt als een virtuele mede-programmeur die ontwikkelaars helpt om sneller betere code te schrijven.

Recent (januari 2024):

- OpenAI:
“We’re launching the GPT Store to help you find useful and popular custom versions of ChatGPT”

Artificiële Intelligentie als onderdeel van 'data science'

1. Generatieve AI

- Large Language Models (LLM's)
- b.v. ChatGPT van OpenAI: geavanceerde chatbot

2. Verwerken en analyseren van grote (meet)datasets

- Foto's, video's, remote sensing data; voorbeeld www.globalwaterwatch.earth

3. 'Data-gedreven' modelbouw

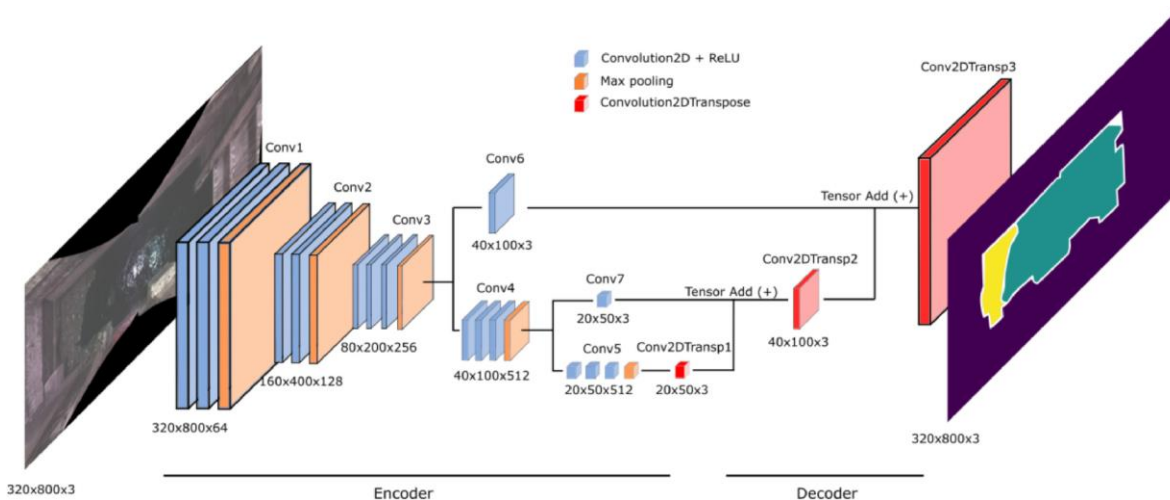
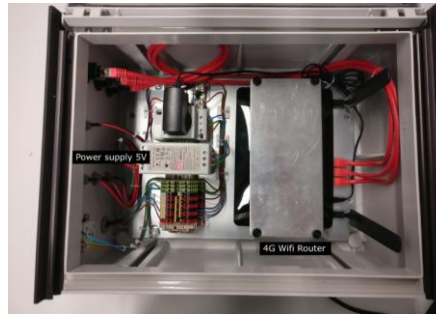
- Classificatie
- Regressie
- Clustering

4. Emulators of 'Surrogate Models'

- Nabootsen van bestaande numerieke modellen

FATracker

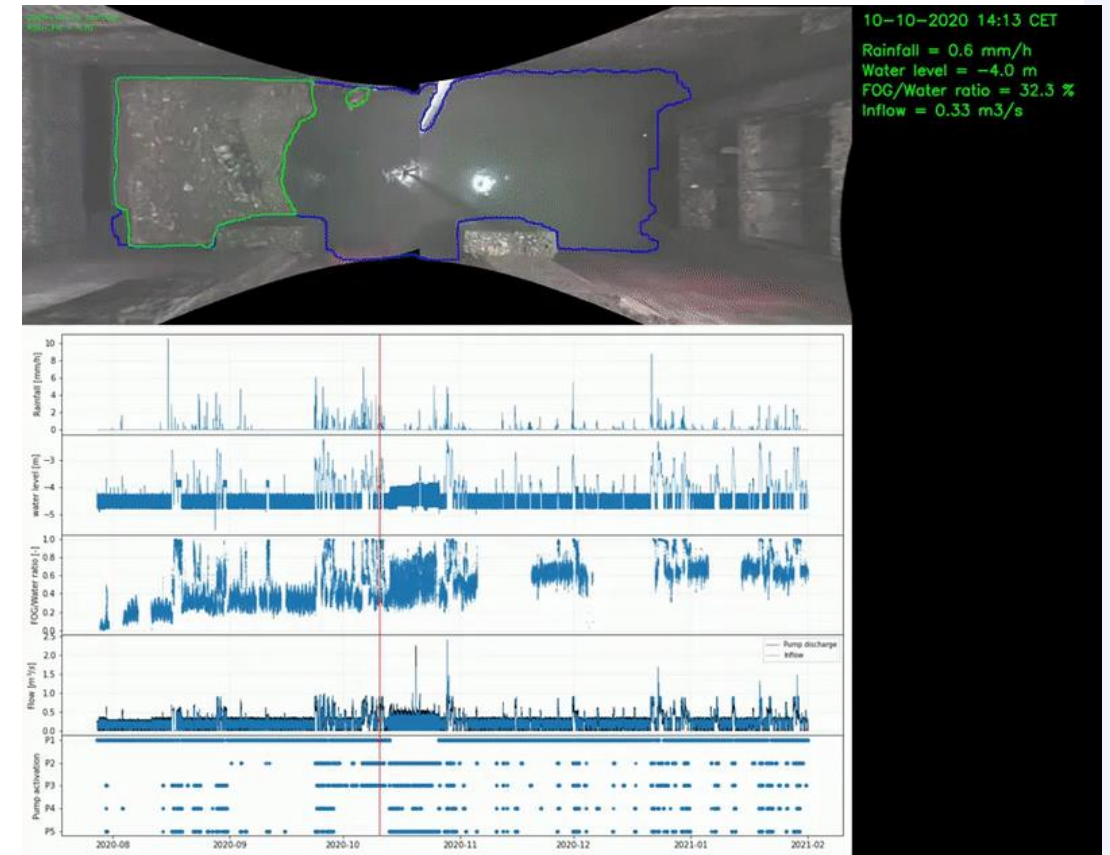
Pumping station fat-oil-grease detection



- Convolutional Neural Network (Encoder-Decoder architecture)
- VGG - Imagenet weights
- Retrained for FOG detection with 124 manually labelled images

Deltares

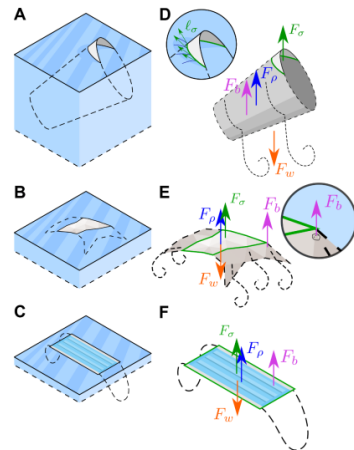
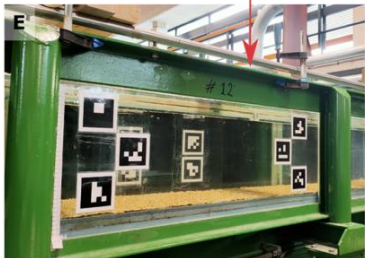
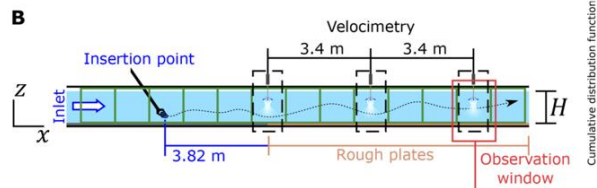
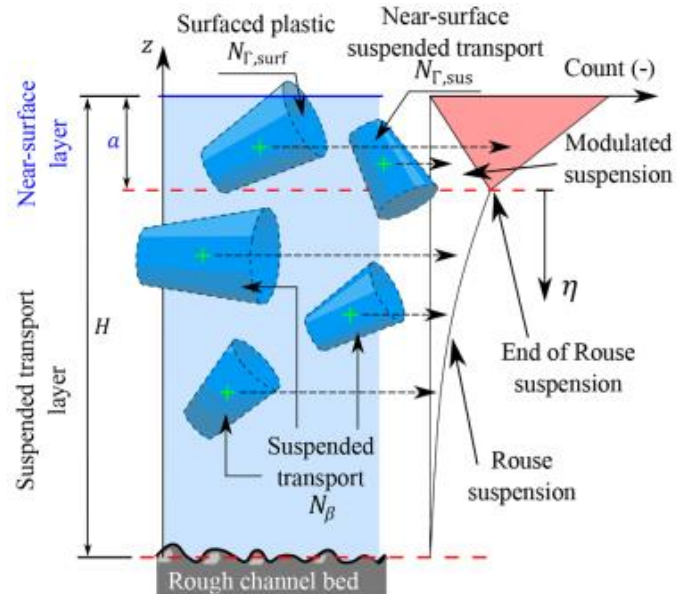
Waterschap Limburg,
Gemeente Rotterdam,
Gemeente Arnhem



Moreno-Rodenas, A. M., Duinmeijer, A., & Clemens, F. H. (2021). Deep-learning based monitoring of FOG layer dynamics in wastewater pumping stations. *Water Research*, 202, 117482.

Computer Vision CNN: Laboratory setups

Hydraulic Laboratory – Automated plastics detection and 3D path reconstruction.
Image processing for targeted object detection and time-space tracking.



Valero, D., Belay, B. S., Moreno-Rodenas, A., Kramer, M., & Franca, M. J. (2022). The key role of surface tension in the transport and quantification of plastic pollution in rivers. *Water Research*, 226, 119078.

Artificiële Intelligentie als onderdeel van 'data science'

1. Generatieve AI
 - Large Language Models (LLM's)
 - b.v. ChatGPT van OpenAI: geavanceerde chatbot
2. AI-gebaseerd verwerken en analyseren van grote (meet)datasets
 - Foto's, video's, remote sensing data
3. 'Data-gedreven' modelbouw
 - Classificatie
 - Regressie
 - Clustering
4. Emulators of 'Surrogate Models'
 - Nabootsen van bestaande numerieke modellen

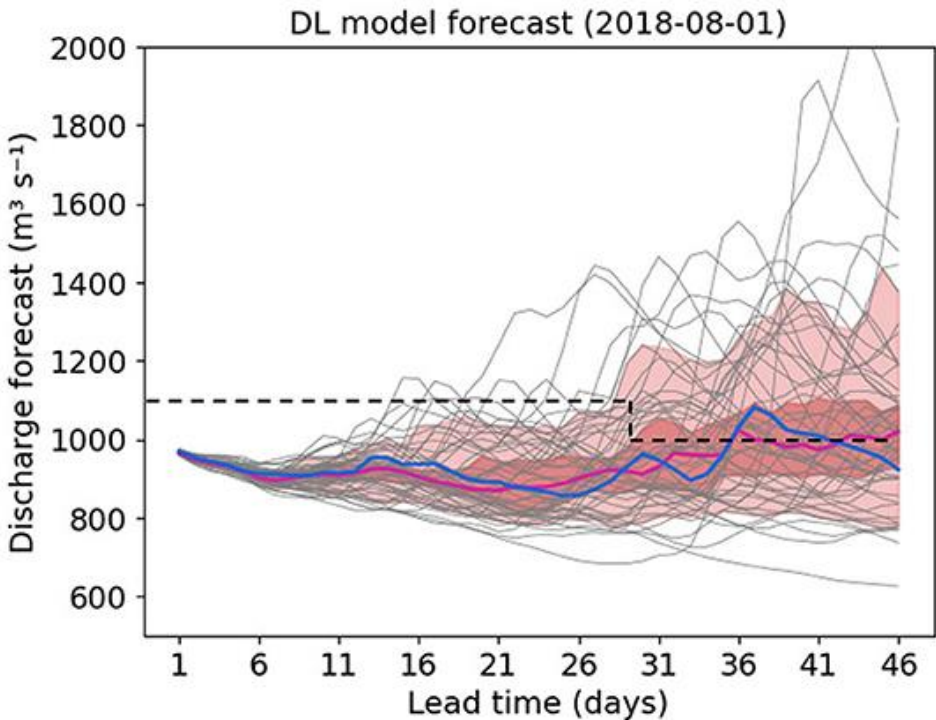
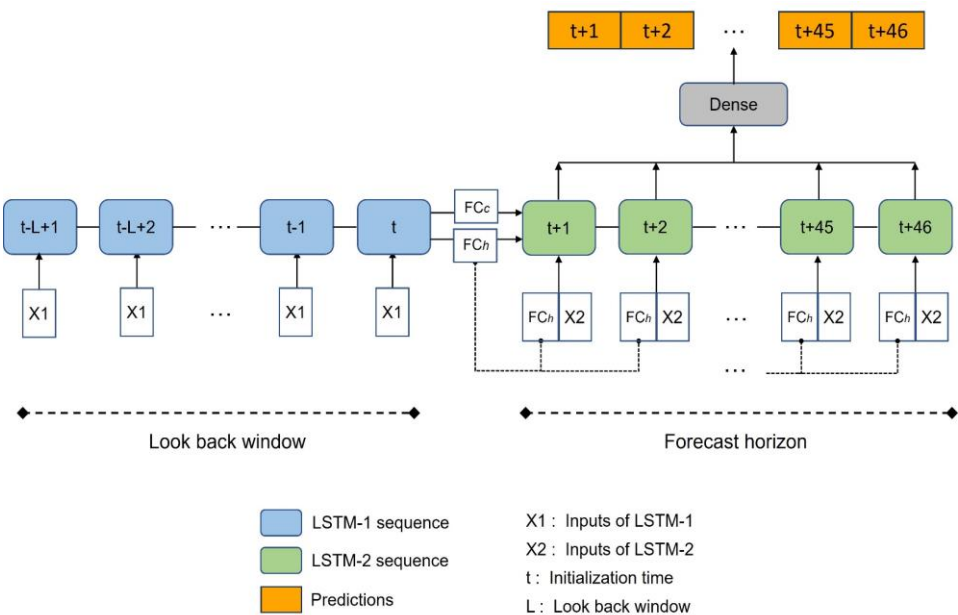
Operational low-flow forecasting model bij Lobith

Deng J., Couasnon A., Dahm R., Hrachowitz M., van Heeringen K.J., Korving H., Weerts A. and Taormina R. (2024). Operational low-flow forecasting using LSTMs. Front. Water 5:1332678. <https://doi.org/10.3389/frwa.2023.1332678>

Een nieuwe, op LSTM gebaseerde modelarchitectuur is ontworpen om zowel historische waarnemingen als voorspelde meteorologische gegevens te gebruiken om de laagwatervoorspelling voor de Rijn bij Lobith uit te voeren op dagbasis met ‘leadtimes’ tot 46 dagen vooruit.



Eight subbasins of the Rhine River basin upstream of Lobith.



Forecast results of the DL model with SEAS5 initialized on 2018-08-01

Afvoerhydrologie

- Long-Short-Term-Memory (**LSTM**)-neurale netwerken doen het al enkele jaren **kwalitatief beter** dan de traditionele Rainfall-Runoff modellen:
 - Een RR-model is een ‘black box model’, er zitten geen deterministische concepten in, neerslag en benedenstroomse afvoer zijn ‘gefit’
 - Een neuraal netwerk is een vorm van ‘multi-dimensionale regressie’ en ‘ontdekt’ daarbij patronen die niet in het gefitte ‘black-box’ model zitten.
 - Recente HESS-artikelen, door Kratzert et al. :
 - [Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets](#)
 - [Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning](#)
- Google + Kratzert:
 - organiseert 22-23 januari het 5^e jaarlijkse online workshop ‘*Flood Forecasting Meets Machine Learning Workshop*’
 - neuralhydrology.github.io

Artificiële Intelligentie als onderdeel van 'data science'

1. Generatieve AI
 - Large Language Models (LLM's)
 - b.v. ChatGPT van OpenAI: geavanceerde chatbot
2. AI-gebaseerd verwerken en analyseren van grote (meet)datasets
 - Foto's, video's, remote sensing data
3. 'Data-gedreven' modelbouw
 - Classificatie
 - Regressie
 - Clustering
4. Emulators of 'Surrogate Models'
 - Nabootsen van bestaande numerieke modellen

Hydrodynamica; emulatie

2D flood modeling:

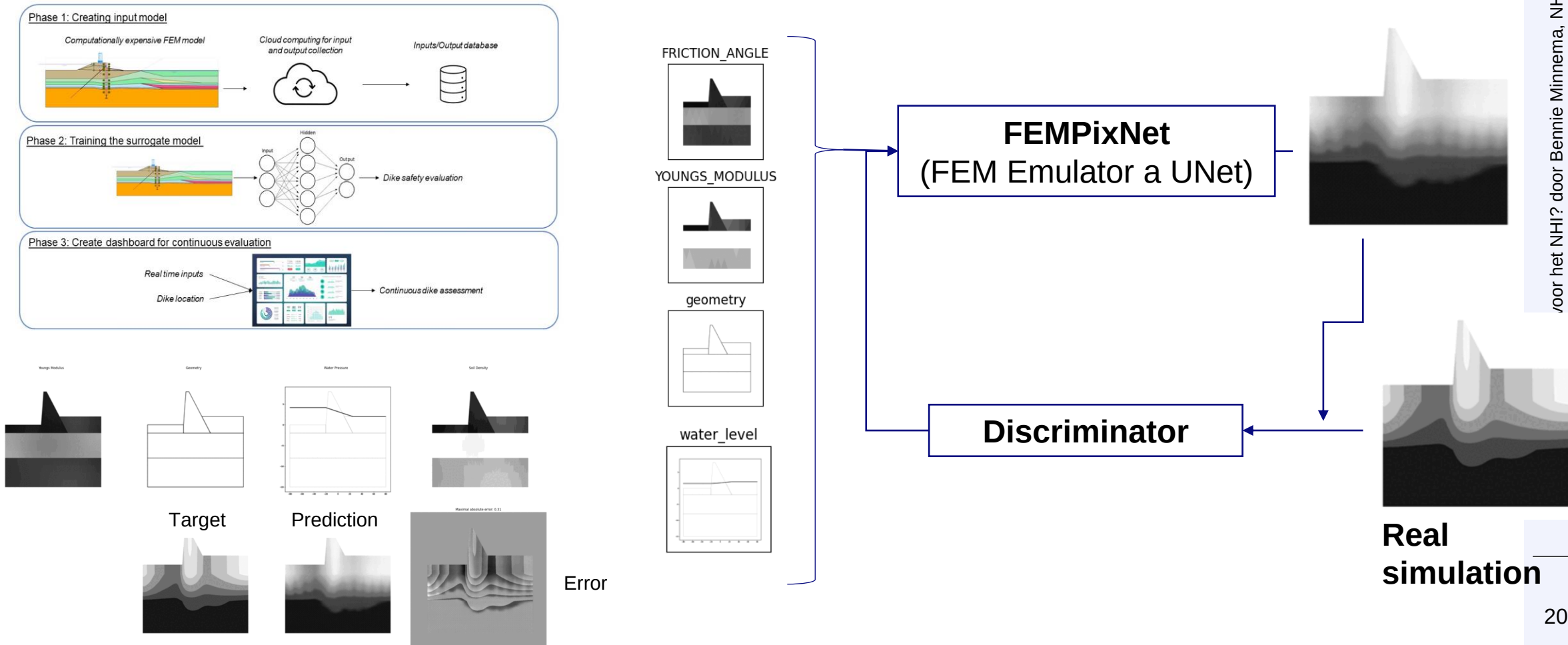
- D-Flow FM, 3Di, HEC-RAS, TUFLOW, etc. bevatten fysieke **deterministische** concepten, voor b.v. ‘overstroming over maaiveld’; echter, lange rekentijden.

Zoeken naar snellere alternatieven:

- Recent symposium (27 september 2023) ‘**Emulating 2D flood modeling**’:
<https://softwaredays.deltares.nl/-/symposium-floods-ai-2023>;
 - er worden al orde van grootte rekentijdwinsten geboekt o.a. gebruik makend van LSTM’s-emulatoren
 - onderzoek lijkt nog maar net op gang te zijn gekomen, er is nog veel te doen.

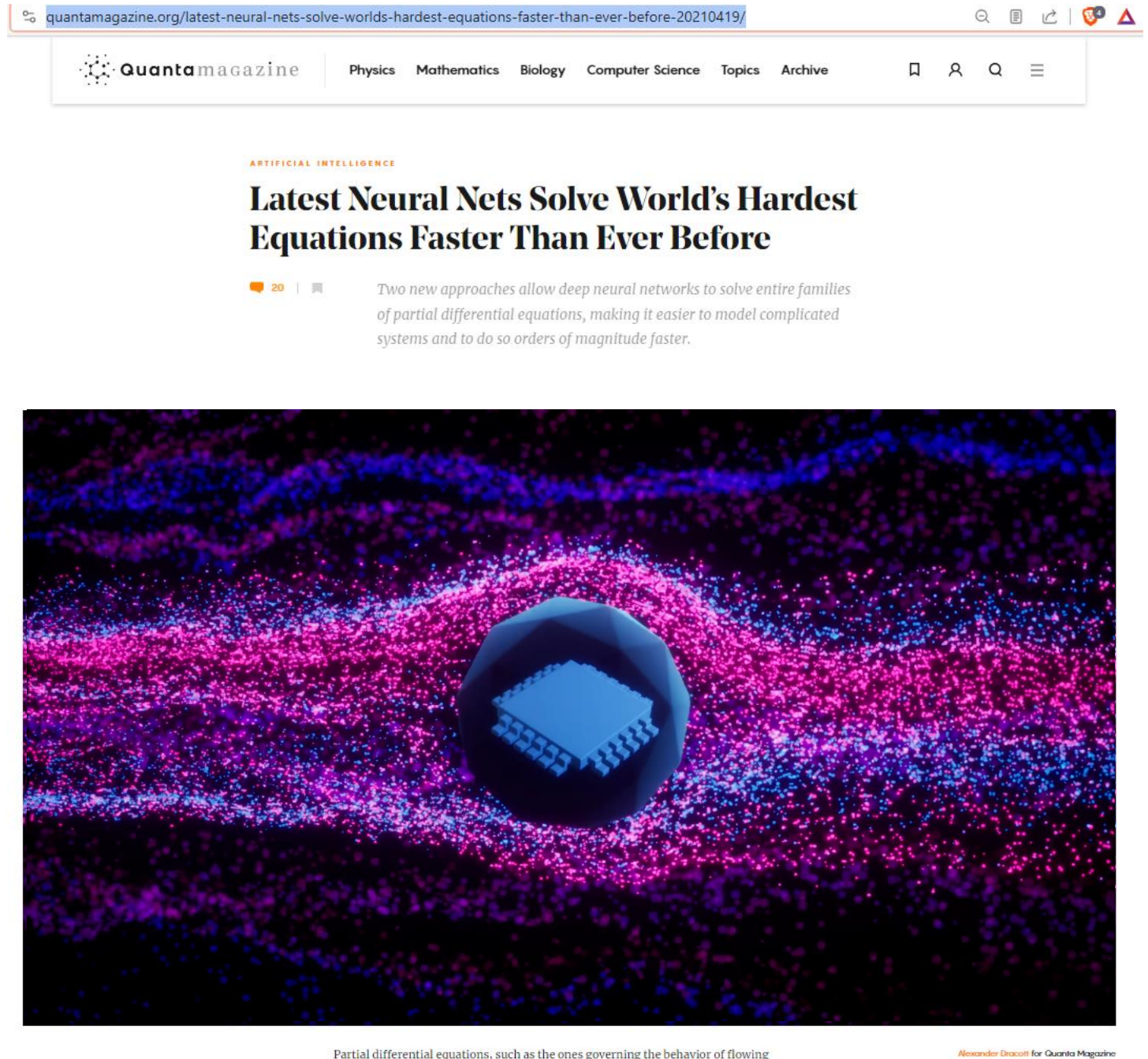
Model Emulation: GAN based emulator of a sheet pile installation (FEMPixNet)

Computationally expensive FEM model of a sheet pile (Kratos + STEM)
5000 simulations (cloud-based as training set)



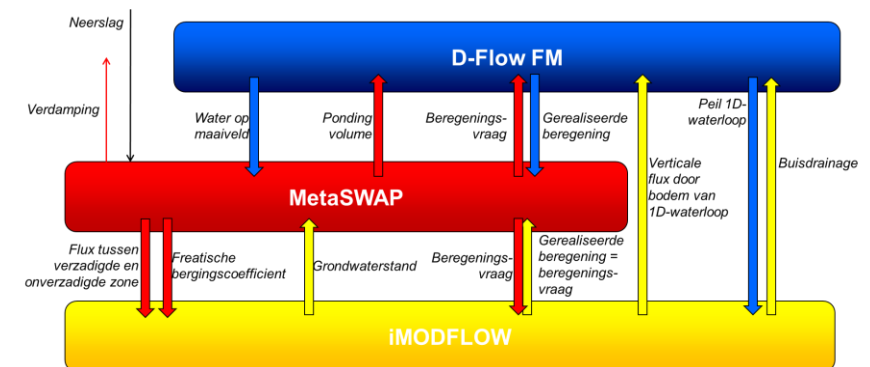
Internationaal ...

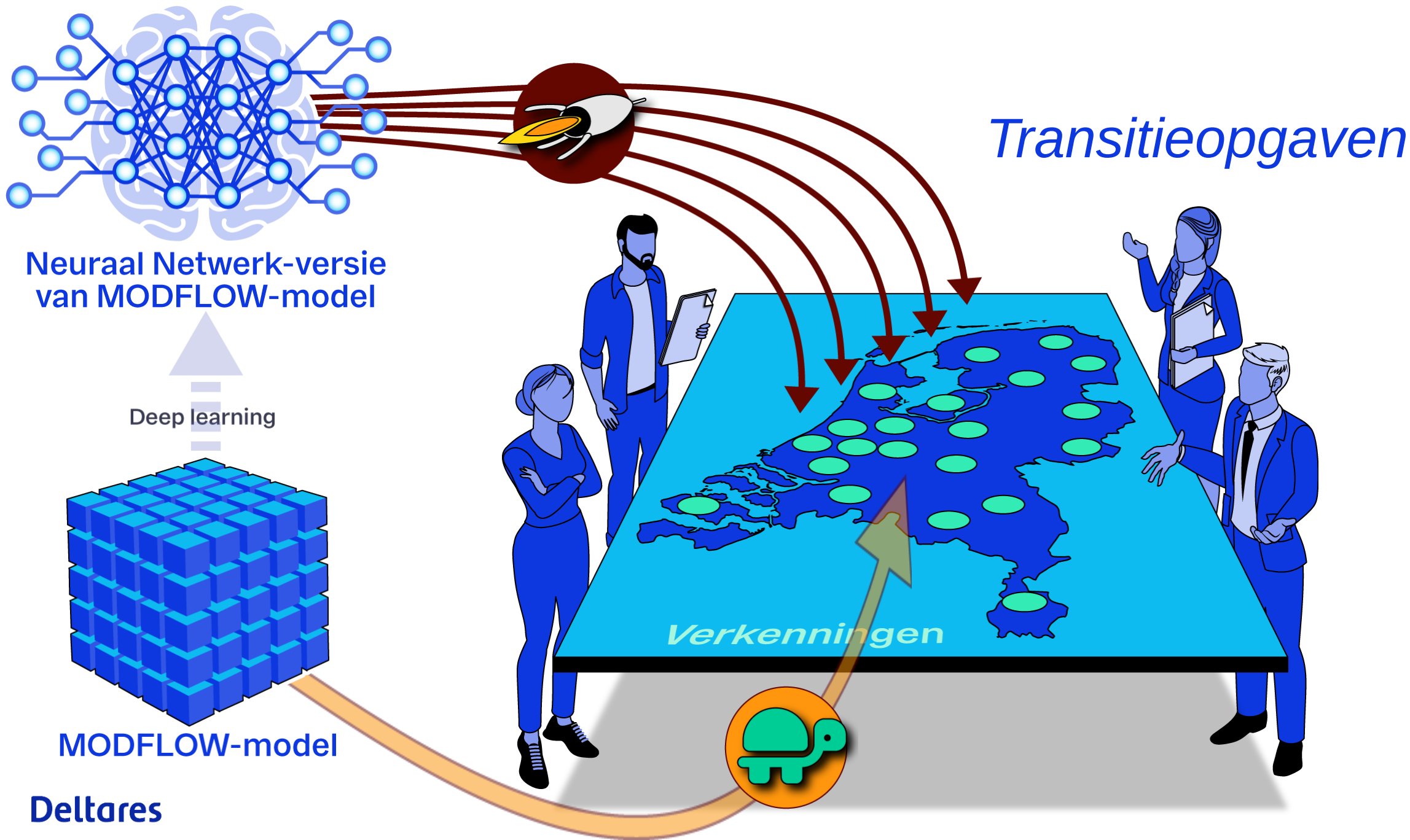
- Algemeen wetenschappelijk onderzoek om partiele differentiaal vergelijkingen veel sneller op te lossen
- O.a. Brown University, USA



Hoe ben ikzelf in de AI-wereld verzeild geraakt?

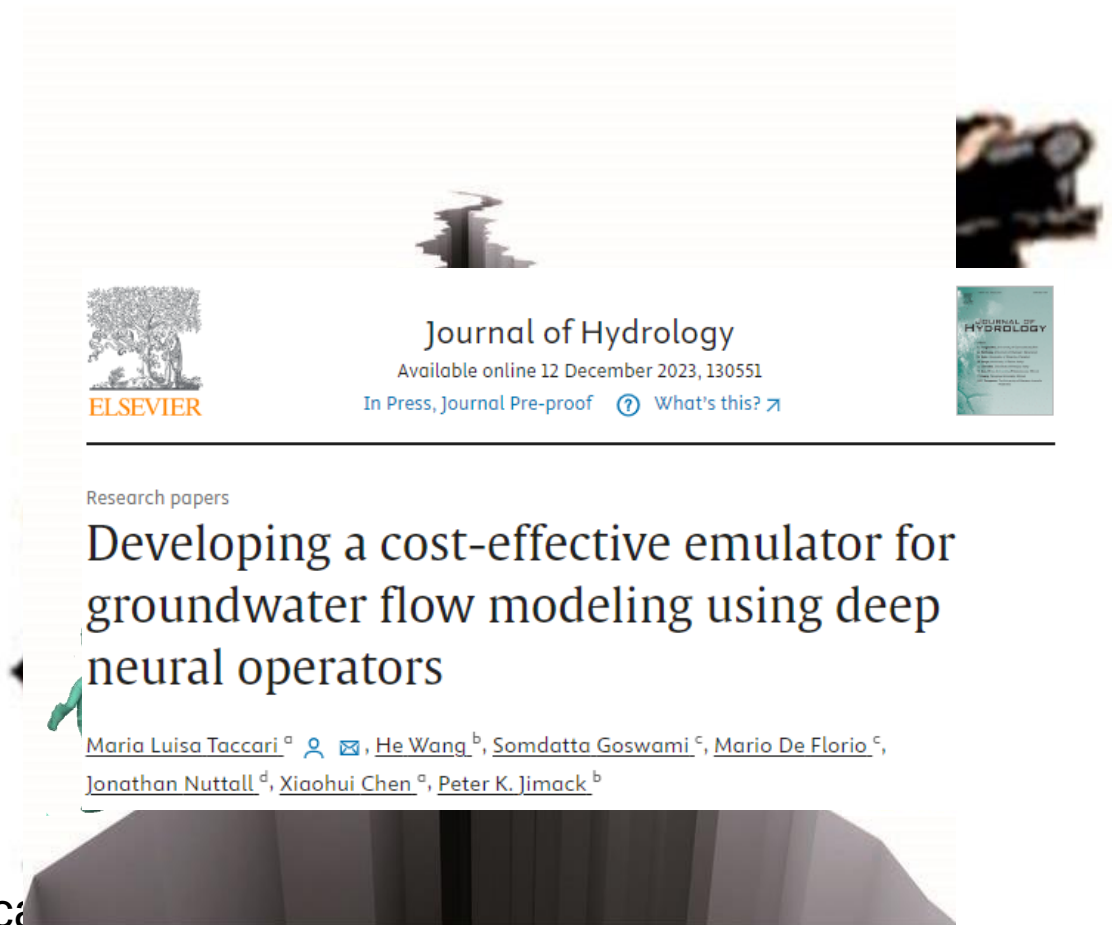
- Sinds ~1990 geohydroloog
- Bouw van grondwatermodellen
 - NAGROM; NAtionaal GRondwater Model, o.b.v. Analytische Elementen Methode (AEM, MLAEM, MVAEM)
 - Strijd tussen AEM, Eindige Elementen Methode, Eindige Differentie Methode => MODFLOW (USGS)
 - Landelijk Hydrologisch Model
 - Parameter optimalisatie (o.a. de 'Representer'-methode van Johan Valstar)
 - Regionale modelbouw o.b.v. iMOD: MIPWA, MORIA, IBRAHYM, HYDROMEDAH, AMIGO, AZURE
- Koppeling grond- en oppervlaktewatermodellen D-HYDRO FM – MetaSWAP - MODFLOW





De techniekontwikkeling; hoe we gestart zijn

- Eind 2017:
Deltares Software Innovation Challenge award
(seed money: ~10 kEuro)
- 2018-2019:
 - Connectie met DSC en Rijkswaterstaat
 - Zoektocht naar funding: RWS ~50 + 40 + SO
 - Samenwerking met Olie & Gas-industrie:
😊 😐 😞: nog een erg gesloten wereld.
 - Gestart met eenvoudige grondwater modellen
 - Brug slaan tussen geohydrologen en wiskundigen
- Medio 2019: eerste resultaten gepresenteerd
op MODFLOW and more conference
- 2020-2024: promotieonderzoek van Maria Luisa Taccari



From: Lucky Luke, De Spookstad, René Goscinny, Belgium

TKI-project: *AI-kennis voor grondwatermodellen*

HydroLogic

HKV
LIJN IN WATER

Witteveen + Bos

Vitens

 Rijkswaterstaat
Ministerie van Infrastructuur en Waterstaat

Deltares

stowa
STICHTING
TOEGEPAST ONDERZOEK WATERBEHEER

 HOOGHEEMRAADSCHAP
DE STICHTSE
RIJNLANDEN

Waterschap NOORDERZIJVEST

 WATERSCHAP
vechtstromen

WATERSCHAP
ZUIDERZEE LAND


Doel:

- Een neurale netwerk-versie maken van bestaande MODFLOW-modellen

zodanig dat:

- Het consistent is met het onderliggende MODFLOW-model

Waarbij:

- getrainde neurale netwerk 100-1000 keer zo snel is als het MODFLOW-model

=>

- Ideaal voor quick-scan doeleinden
- Ideaal voor parameter optimalisaties en Monte Carlo simulaties
- 'iPAD'-versies van grondwatermodellen

TKI-project *AI-kennis voor grondwaterverkenningen*

1^e jaar:

- Testen verschillende typen neurale netwerken (UNet, GAN, CNN, FNO, DeepONet, LSTM) voor kleine(re) (hypothetische) testmodellen.

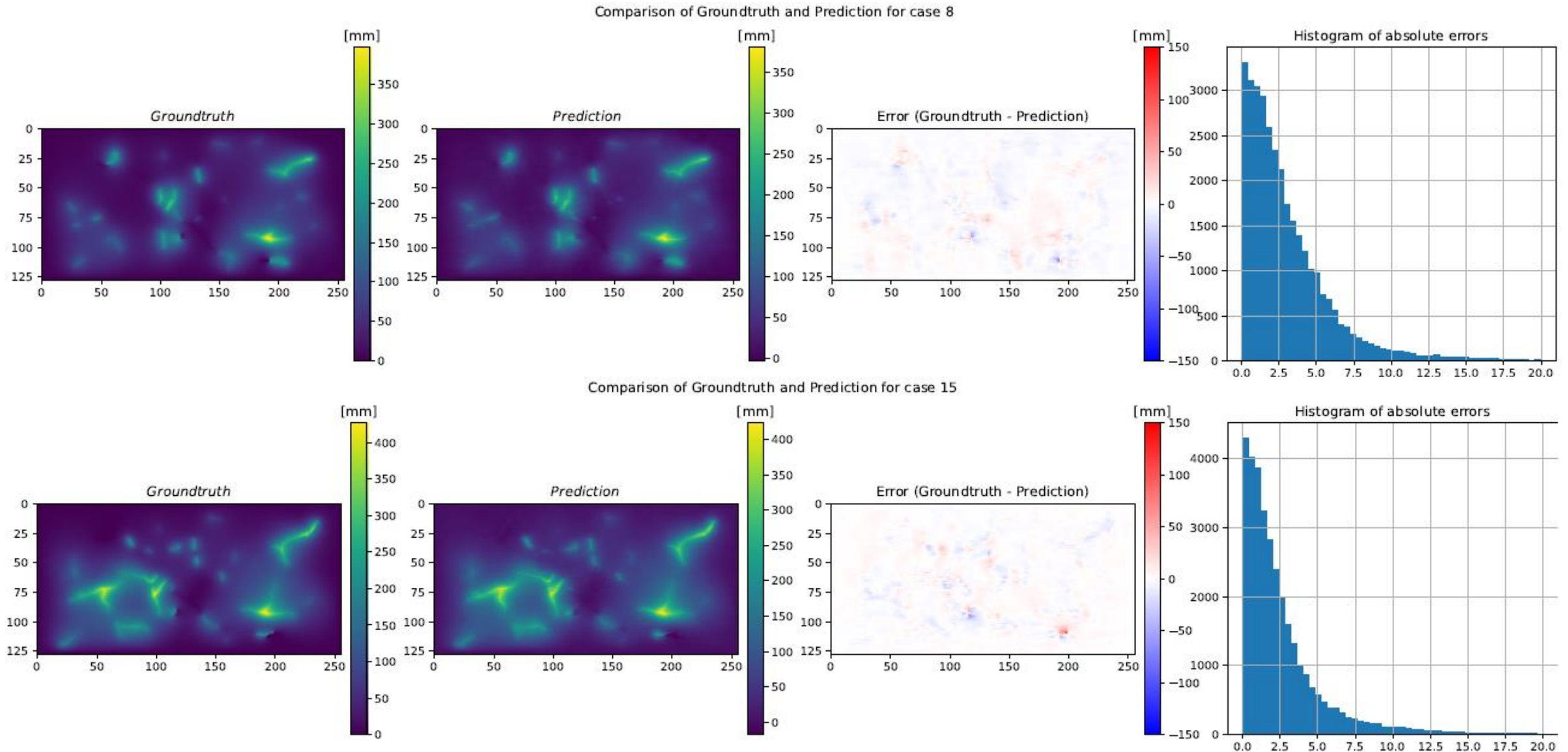
2^e jaar:

- Opschalen naar 'real-world' casussen

Meer info:

<https://tkideltatechnologie.invoermodule.nl/project/ai-kennis-voor-grondwaterverkenningen/>

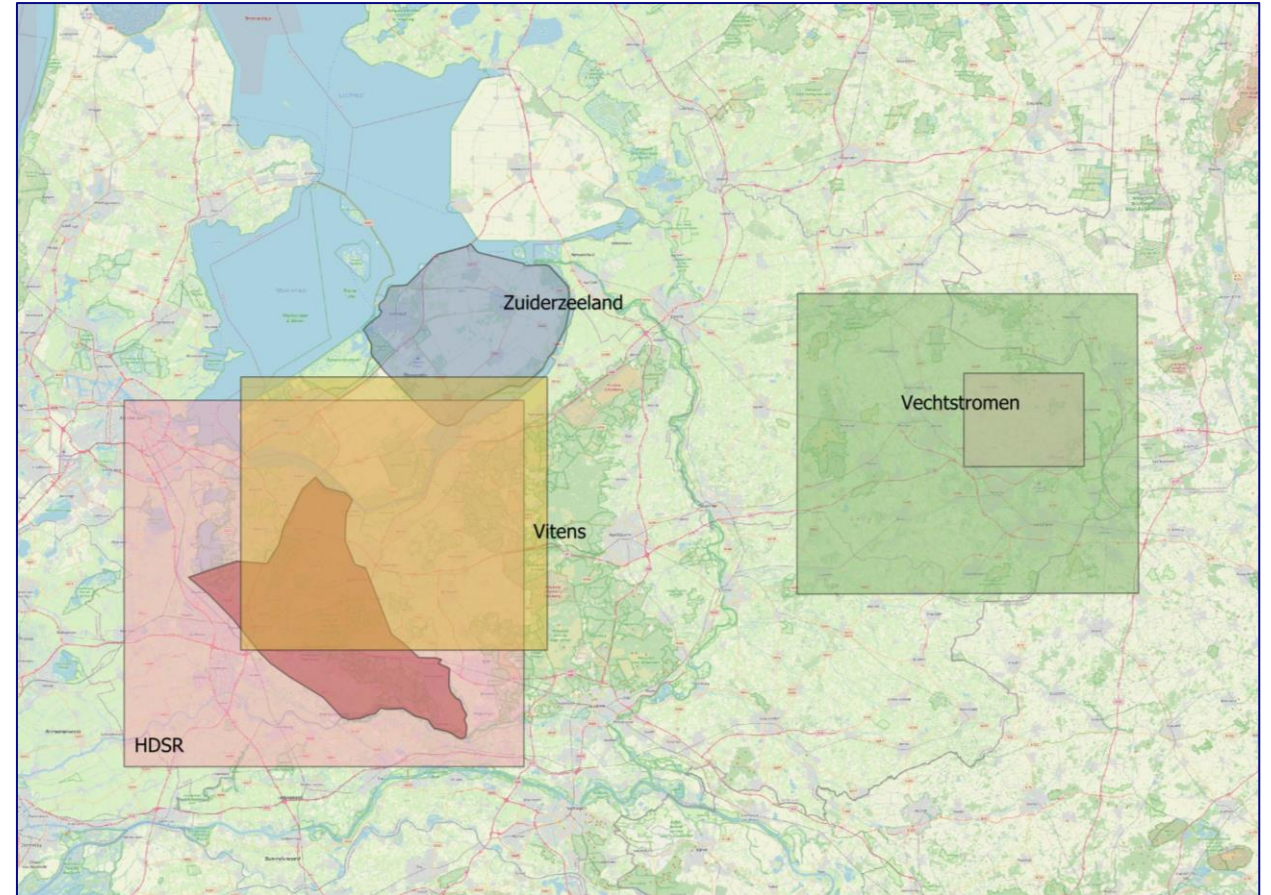
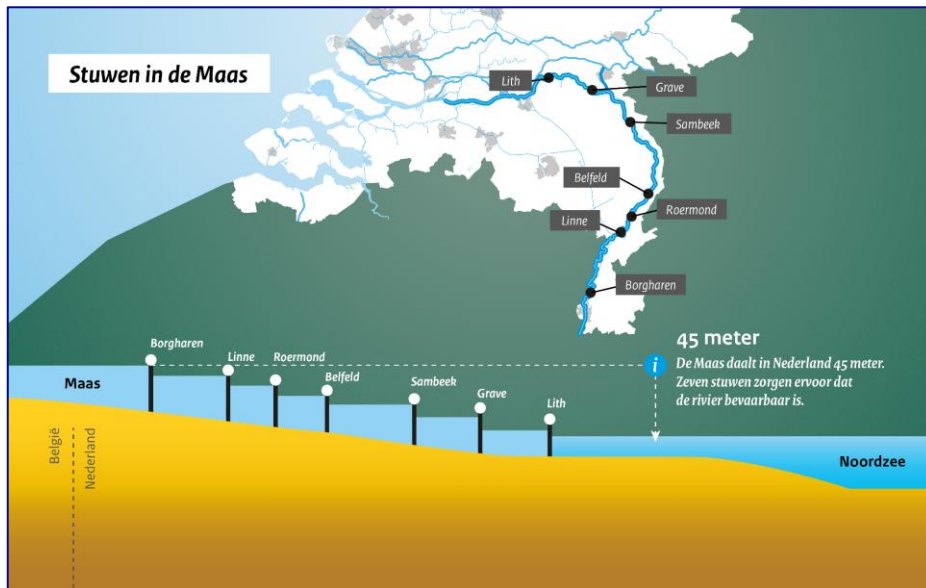
Voorbeeld: Fourier Neural Operator (FNO); effecten op GLG



Deltares

TKI-project *AI-kennis voor grondwaterverkenningen*

Real-world casussen:



TKI-project *AI-kennis voor grondwaterverkenningen*

- Einddoel:
dat we in staat zijn razendsnelle emulators te maken van bestaande MODFLOW-modellen.
- Maar ook:
generieke aspecten kunnen hergebruiken voor belendende domeinen, bijvoorbeeld bij het simuleren van riviermorfologie.

(nabije) toekomstplannen

1. Enabling Technologies team, Deltares

- maakt op dit moment een update van het overzicht wie binnen Deltares met welke AI-technieken bezig zijn => kennisuitwisseling.
- voorstel: dit overzicht uitbreiden binnen NHI-kader.

2. AI-grondwater symposium dag, in week 47 (18-22 november 2024, onderdeel van Deltares Software Days).

Oproep:

- Alvast na te denken over een AI en grondwater gerelateerde presentaties.
- Alvast nadenken over het opzetten van AI en (grond)water-gerelateerde vervolgprojecten.

Dank u voor uw aandacht!

Voor tips & feedback, email naar Bennie.Minnema@deltares.nl

- **Artificial intelligence:** The field of computer science focused on intelligent computer programs that can sense, reason, act, and adapt.
- **Machine learning:** A subset of AI in which algorithms can improve in performance over time when exposed to more data.
- **Neural network:** A series of algorithms used as a process in machine learning that can recognize patterns and relationships in large quantities of data. Neural networks use a logic structure inspired by the human brain and are the foundation for deep learning algorithms.
- **Deep learning:** A subset of machine learning in which multilayered neural networks learn from vast amounts of data.

Kunstmatige intelligentie: het vakgebied van de informatica dat zich richt op intelligente computerprogramma's die kunnen waarnemen, redeneren, handelen en zich kunnen aanpassen.

Machine learning: een subset van AI waarbij algoritmen in de loop van de tijd hun prestaties kunnen verbeteren als ze aan meer gegevens worden blootgesteld.

Neuraal netwerk: een reeks algoritmen die worden gebruikt als proces bij machinaal leren en die patronen en relaties in grote hoeveelheden gegevens kunnen herkennen. Neurale netwerken gebruiken een logische structuur geïnspireerd door het menselijk brein en vormen de basis voor deep learning-algoritmen.

Deep learning: een subset van machine learning waarbij meerlaagse neurale netwerken leren van enorme hoeveelheden gegevens