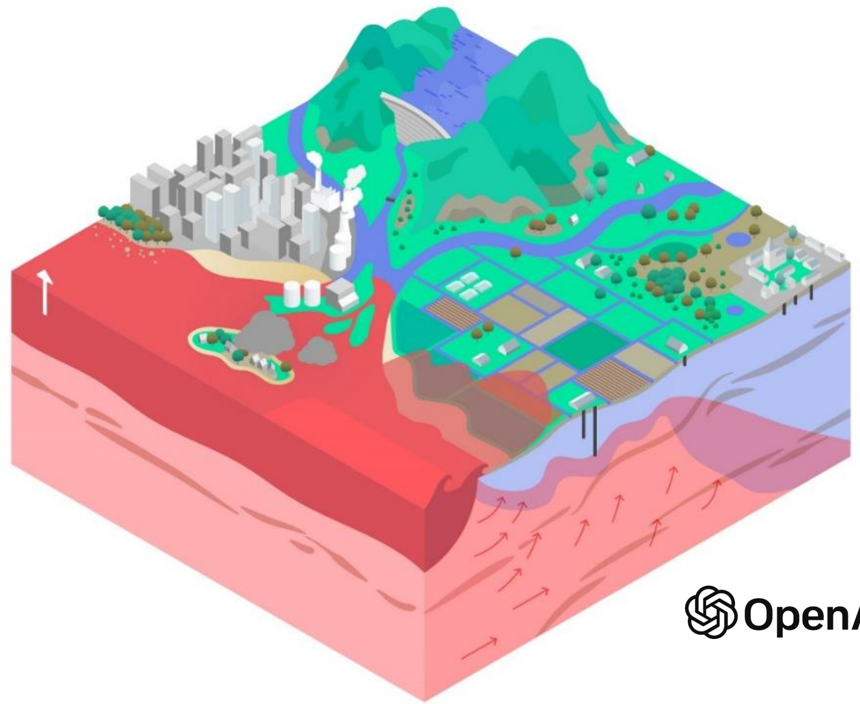
Session HS8.1.8

EGU24-10924

EGU General Assembly 2024



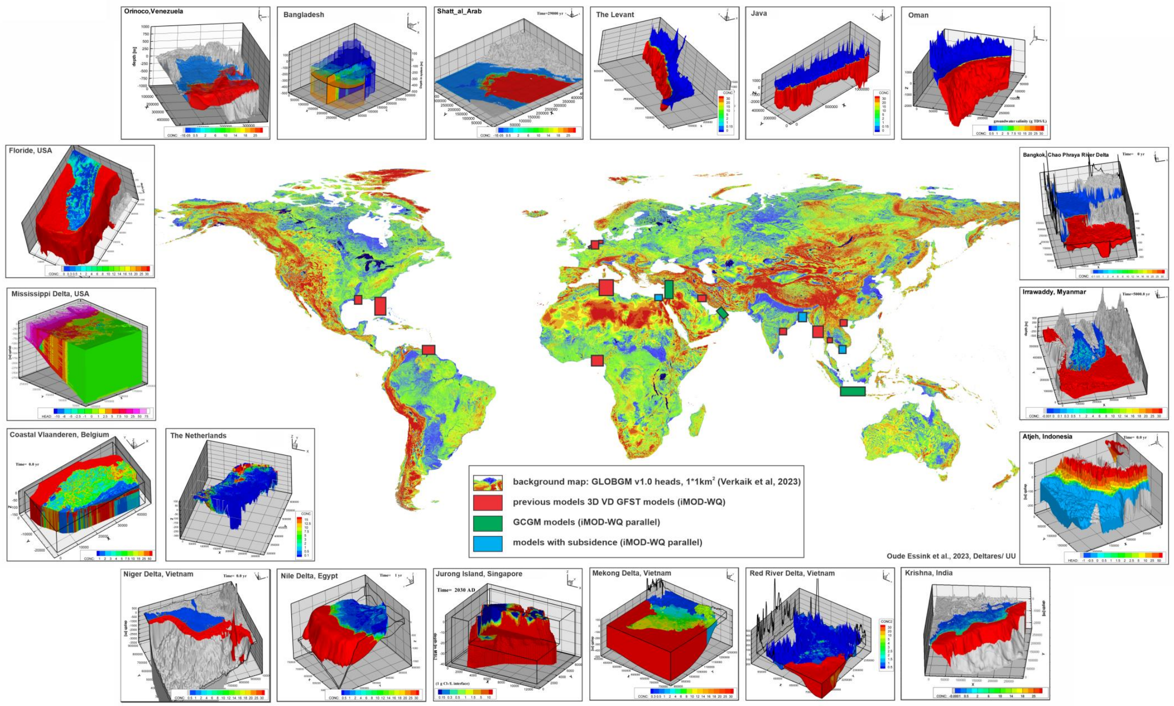
Tesseract OCR

PyMuPDF

OpenAI

Microsoft Azure

Building large-scale 3D coastal groundwater models with iMOD-WQ and global datasets



Daniel Zamrsky[1], **Gualbert H.P. Oude Essink**[2,1], Jude King[2], Wahdan Achmad Syaehuddin[3], Mitchel Kwint[4], Marc F.P. Bierkens[1,2]

1 Utrecht University

2 Deltares

3 Wageningen University & Research

4 The Hague University of Applied Sciences

Why this coastal groundwater model initiative?

- Coastal fresh groundwater is **main water resource** for ~**50%** of the world population in the coastal zone
- Groundwater is important for **drinking water, agriculture**, **industry**, as well as **ecosystems & river baseflows**
- **Right now**, fresh groundwater resources are threatened by **excessive pumping**, and **reduced replenishments**
- Projected climate change impacts, inducing **sea-level rise**, will worsen this situation
- We need quantified **storylines** on fresh groundwater availability under stress in **data-poor** coastal zones
- These storylines should be **linked** to **droughts, land subsidence, flooding, (human) health and biodiversity**

Image credit: wikipedia

50% *of our drinking water is fresh groundwater*

Image credit: USGS

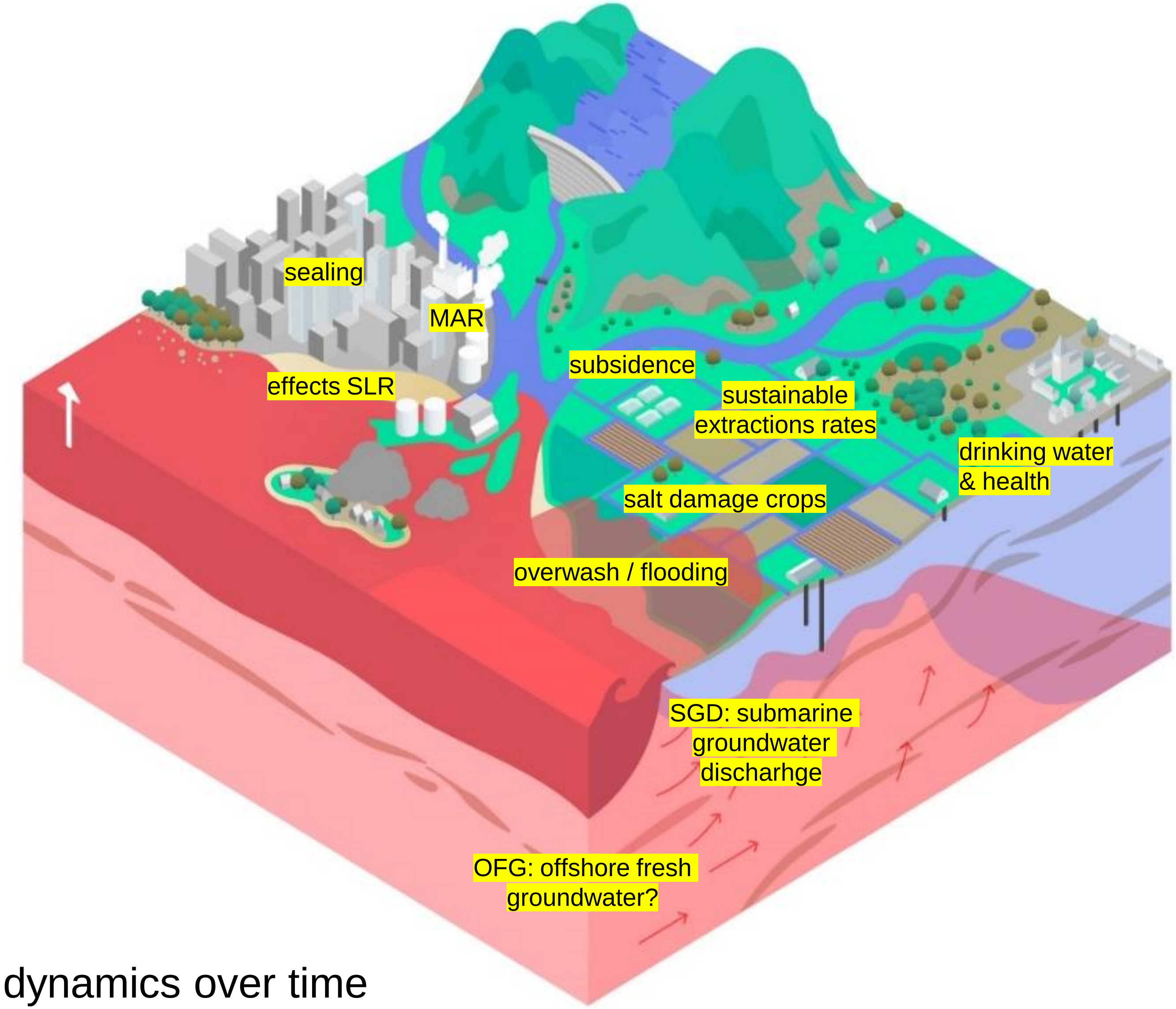
Image credit: Piyaset/Shutterstock.com

38% *of irrigable area is irrigated with fresh groundwater*

Image credit India Water Portal

Applications, components and insights large-scale coastal groundwater model

- Components:
  - groundwater quantity
  - groundwater salinity
  - subsidence (into 2024)
  - heat transport (later, >2024)
  - groundwater quality (later, >>2024)



- Insights:
  - into (supra) - regional coastal groundwater dynamics over time
  - into understanding current and future state of transboundary fresh groundwater resources
  - into identifying potential hotspots for fresh groundwater shortages

Characteristics large-scale coastal groundwater models LCGMs(1km^2 scale)

1. Models can cover areas **>> 10.000 km^2**
2. Typically **cell sizes** of **1*1km^2**
3. **Multiple model layers** (to properly represent groundwater salinity and subsidence)
4. Our LCGM building tool use **open-source tools like Python**
5. **Global datasets** are used, providing 1st-order approximations of groundwater conditions in data scarce regions
6. **High-performance computing** opens up new possibilities
7. **Parallelization of SEAWAT** (iMOD-WQ): important breakthrough in speeding up variable-density groundwater flow and salt transport modeling
8. **Simulation** groundwater salinity dynamics over **full glacial-interglacial cycle** (e.g. 125 ka).

Characteristics large-scale coastal groundwater models LCGMs(1km^2 scale)

9. **HydroBASINS global-watershed-boundaries** dataset are used to delineate the boundaries
10. **Offshore continental shelves** are also covered (manually outlined and added to the selected HydroBASINS).
11. **Top elevation** derived from a **global DEM dataset (GEBCO)**
12. **Bottom elevation** estimated by:
    a. the bottom of the unconsolidated sediment formations
    b. sedimentary rock formations (limited to siliclastic lithology)
13. When **local hydrogeological input data** is available (e.g. borelogs, groundwater salinity, extractions), tools like **GEMPY** are used to **improve the LCGMs**

Planning large-scale coastal groundwater model building tool

Phase 1: Tool development, demonstration

Phase 2: Global modelling and scenarios

Key
- External data / mining
- Tool development
- Products

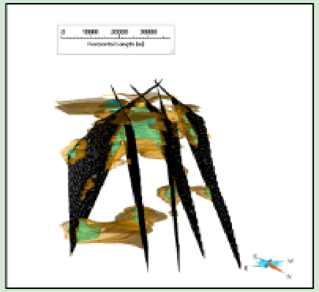
2024 salinity

In-situ data (water samples, airborne EM etc)

Global 2D Salinity profiles (Zamrsky)

Global, rapid, automated 3D interpolation

3D Salinity (initial concentration)



2024/2025: Global geology/modelling

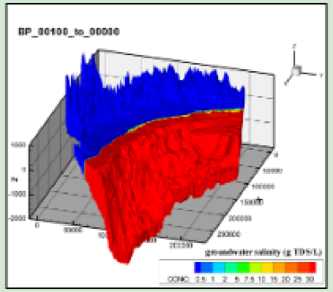
In-situ data (boreholes/seismics etc)

Global 2D lithology profiles (Zamrsky, aquifer base)

Flexible 3D model creation (with or without in-situ data)

3D Geology (hydraulic properties)

Demonstrate framework with iMOD-WQ parallel

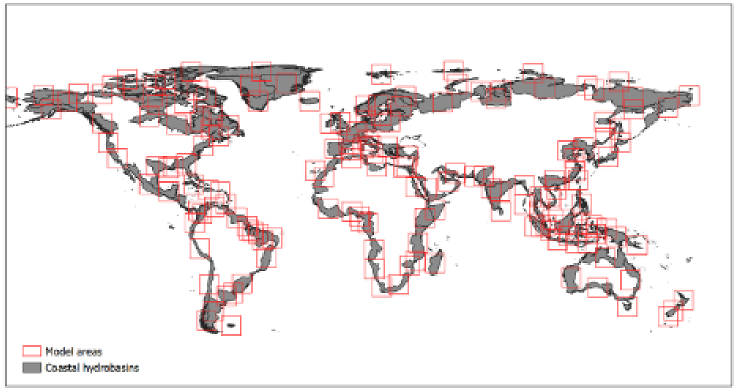


2025/2026: Calibrated global salinity models

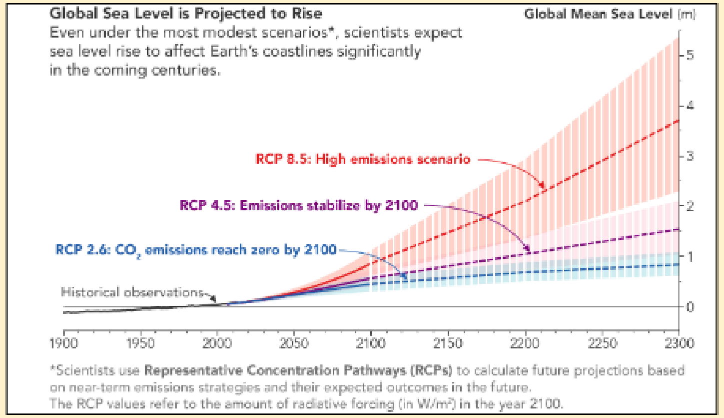
Globally available calibration data (from data mining/ML)

3D global modelling toolbox (COSCAT regions)

3D global salinity models



2025/2026: Climate scenarios & impacts

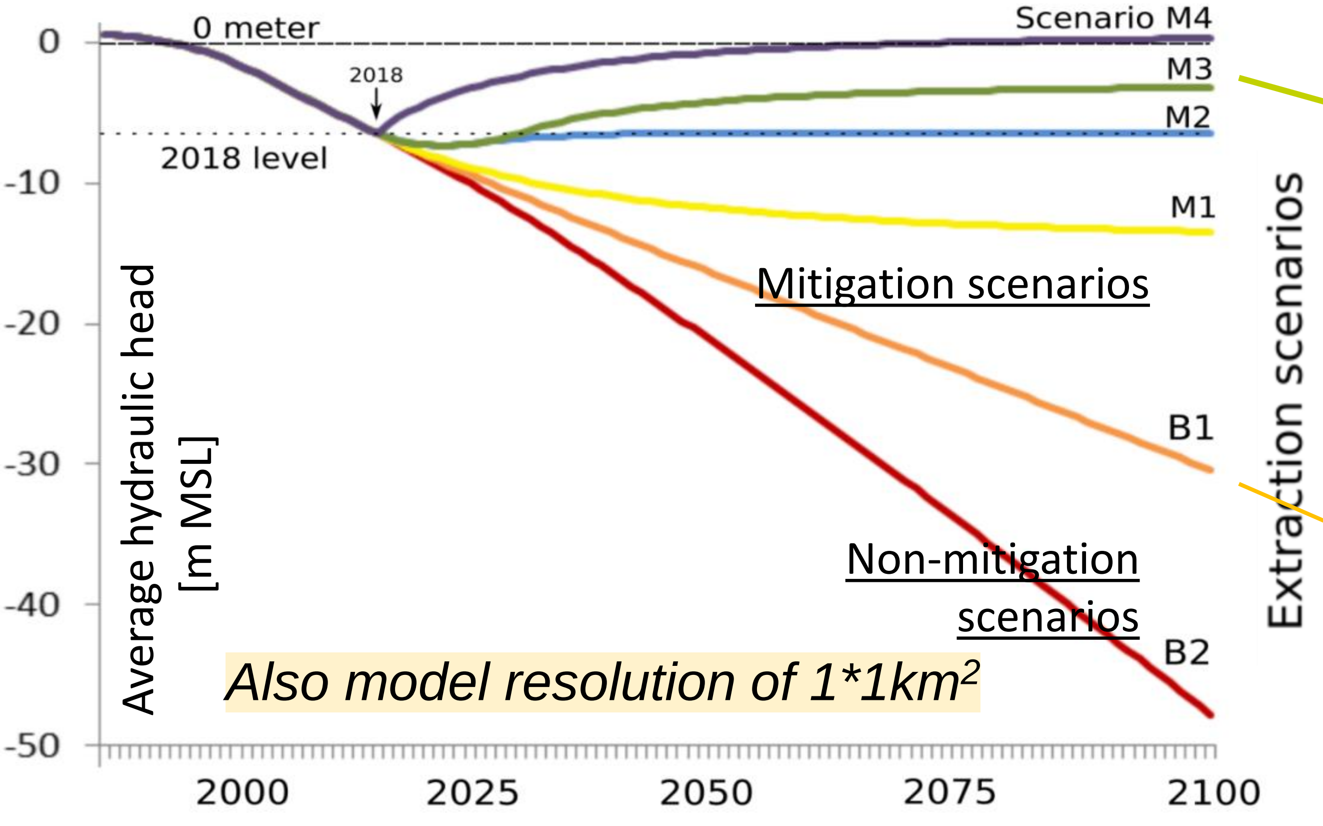


Source: NASA Earth Observatory

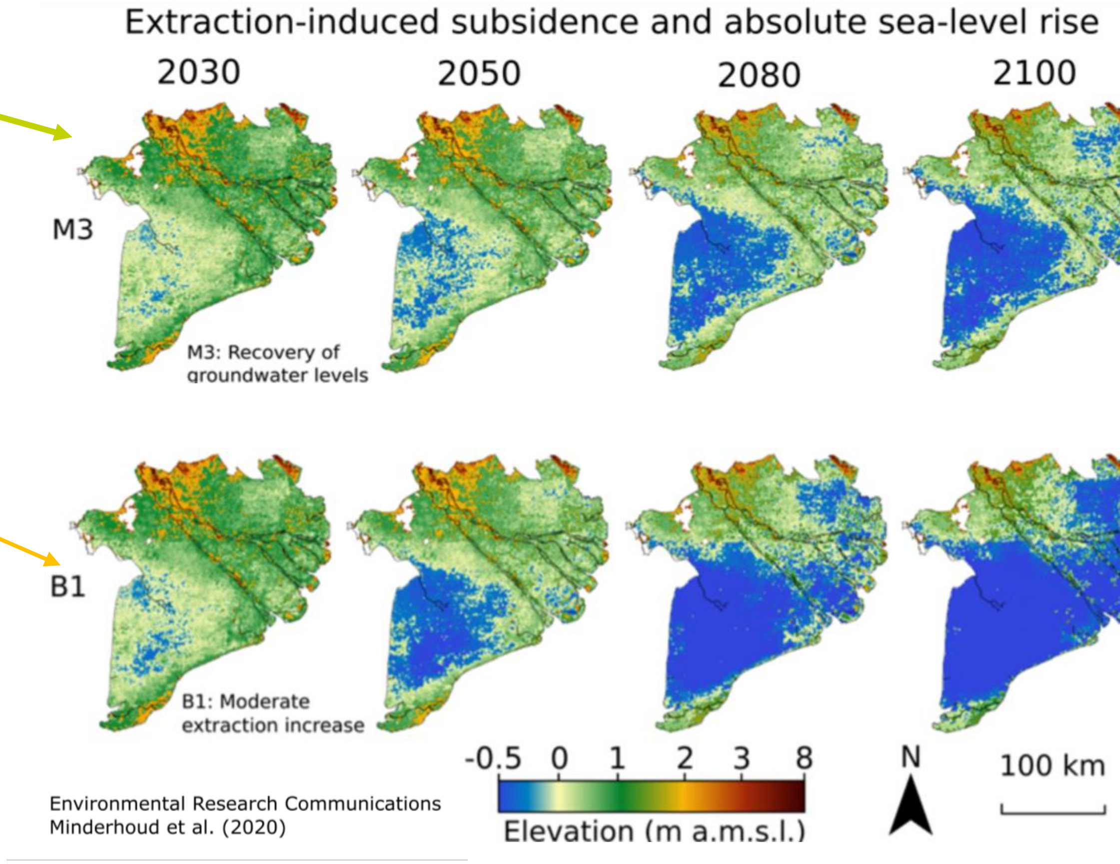
Building large-scale 3D coastal groundwater models with iMOD-WQ and global datasets

Example storyline: Pathways to demonstrate the future Mekong delta: linking groundwater extraction ➡ subsidence ➡ increased flood risk

Scenarios of future groundwater extraction pathways



Future elevation of the delta

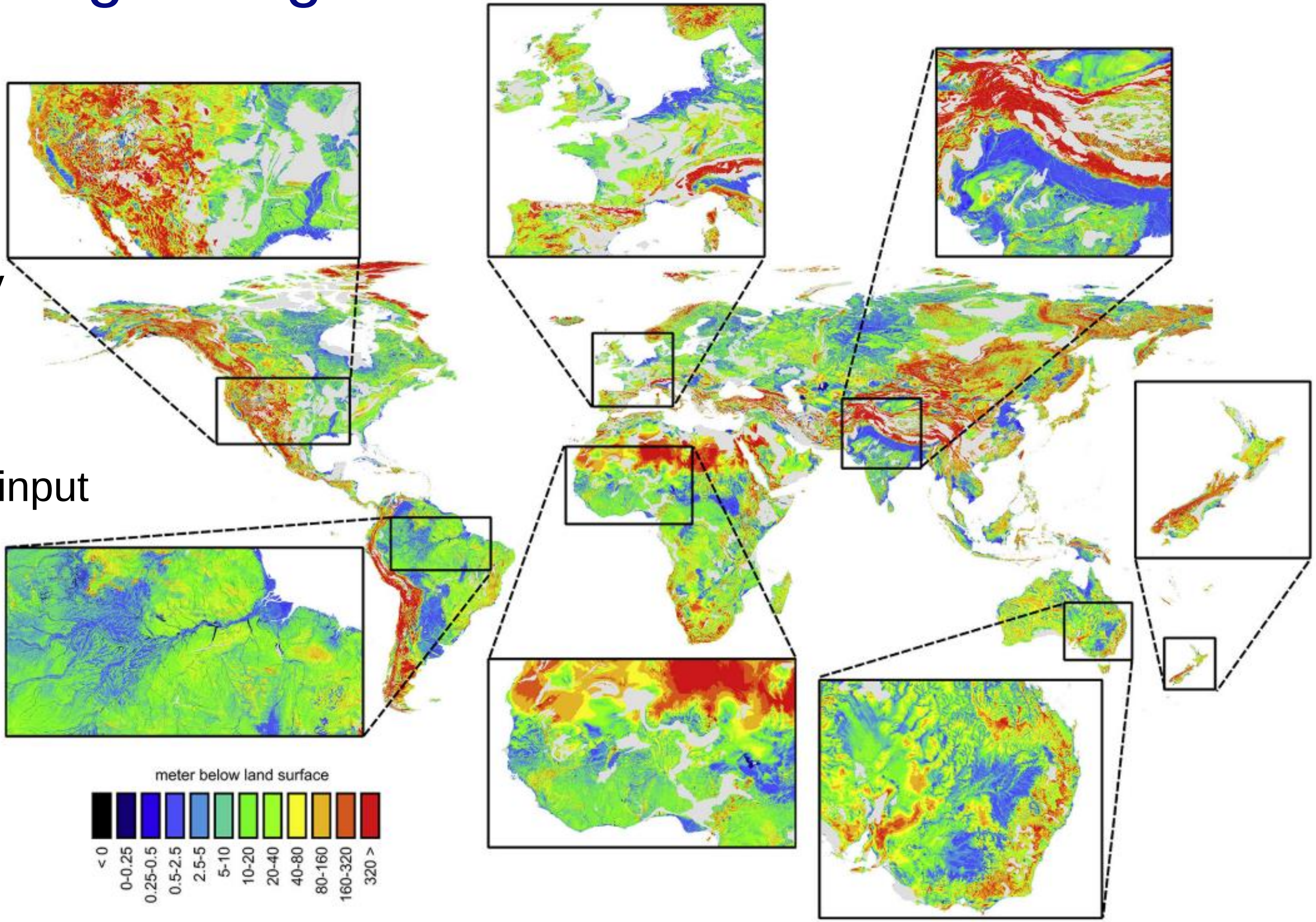


Building large-scale 3D coastal groundwater models with iMOD-WQ and global datasets

Among others based of this research, 'Decree 167' has been implemented in Vietnam in the **Law of Water Resources**: develop and implement zoning plans to **restrict groundwater overexploitation**

Example 1: 1*1km^2 global grondwater model

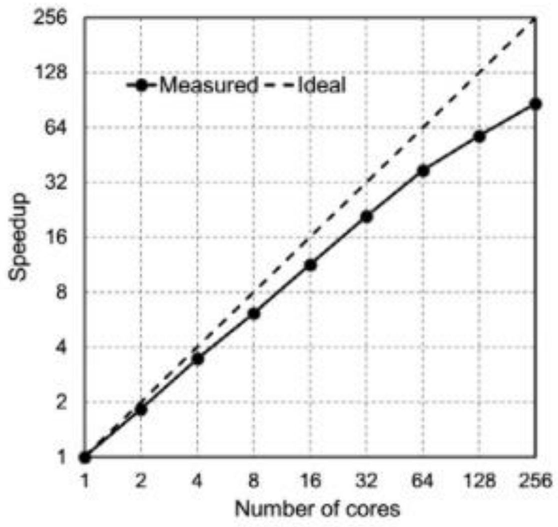
Components:
- quantitative groundwater only
- 278 million active cells
- two model layers
- simulating period 1958–2015
- daily time steps and monthly input
- 12 nodes, 384 cores
- Snellius supercomputer
- maximum 16 hr simu time!



Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H. P., Lin, H. X., & Bierkens, M. F. P. (2024). GLOBGM v1.0: a parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW global-scale groundwater model. *Geoscientific Model Development*, *17*(275–300). https://doi.org/10.5194/gmd-17-275-2024

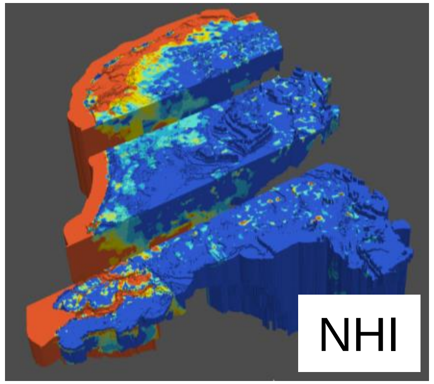
Example 2: parallel computing plus smart model parameters

- Split into (tens of) partitions, leading to a significant reduction in computation time
- Speed-ups of at least 10 up to 100 times, depending on cores, solver iterations and data exchange efficiencies

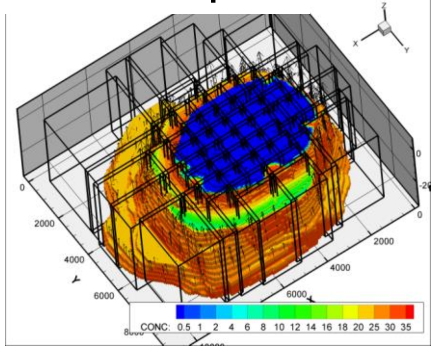


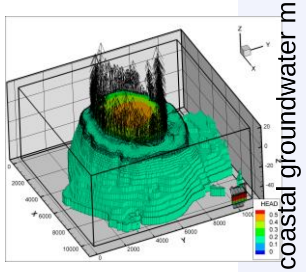
Three examples:

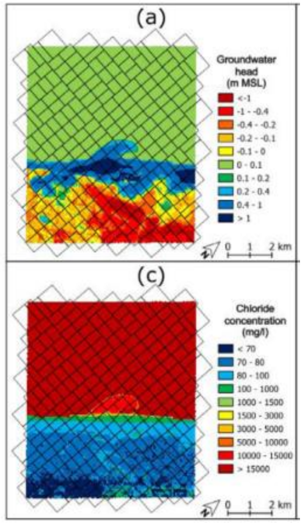
1. Sand Engine: from 1hr 47min 55sec -> 2min 40sec: **40***
2. NHI fresh-salt: from ~30 days to ~2days: **15***
3. Island Japan: from 5d0h36m to 5m59s: **1209***

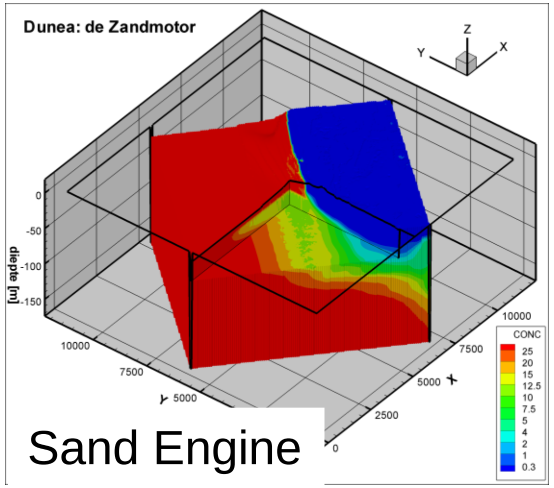


Island Japan

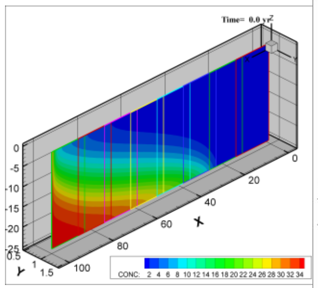


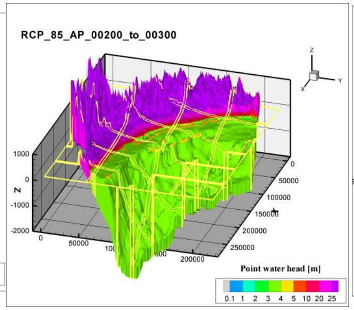


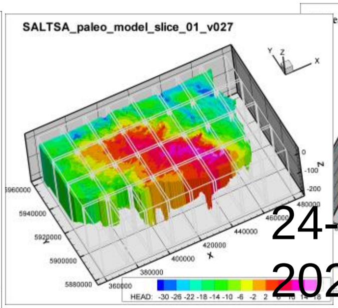


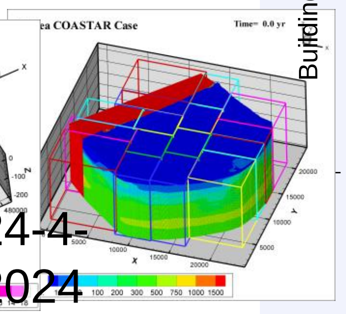


Verkaik, J. et al., 2021. Adv. Water Resources





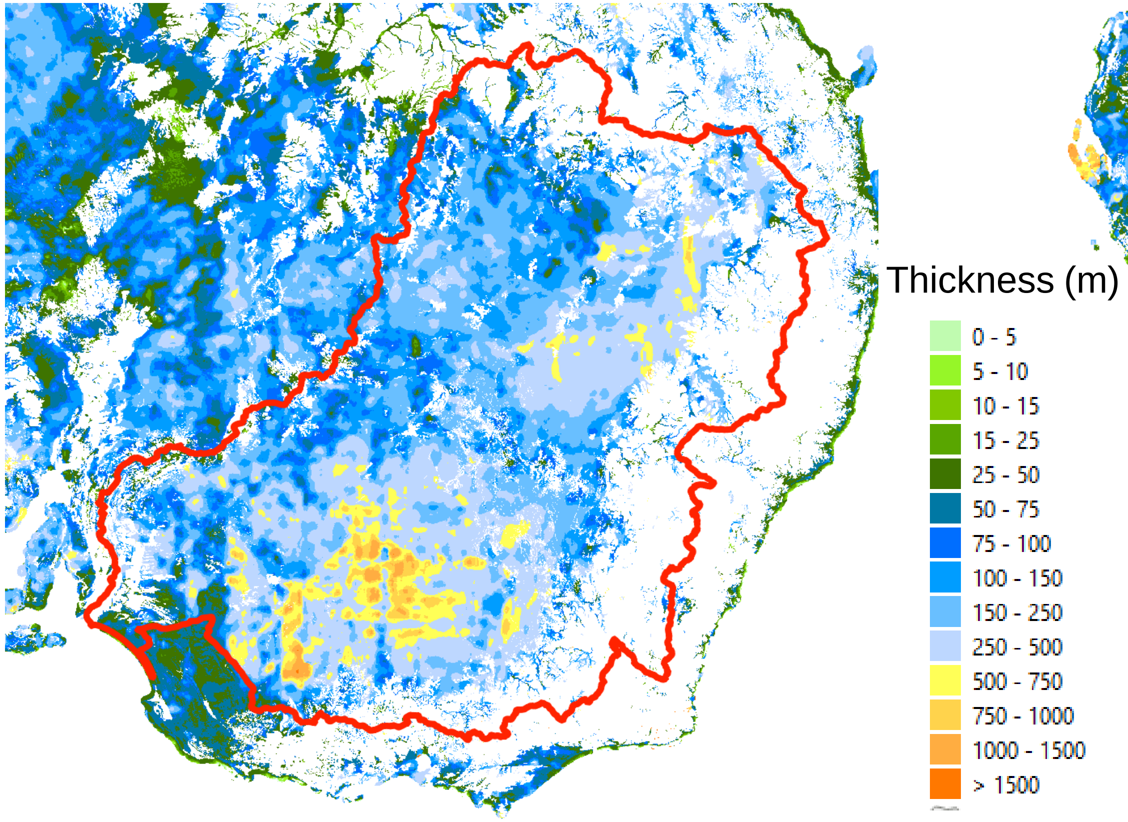


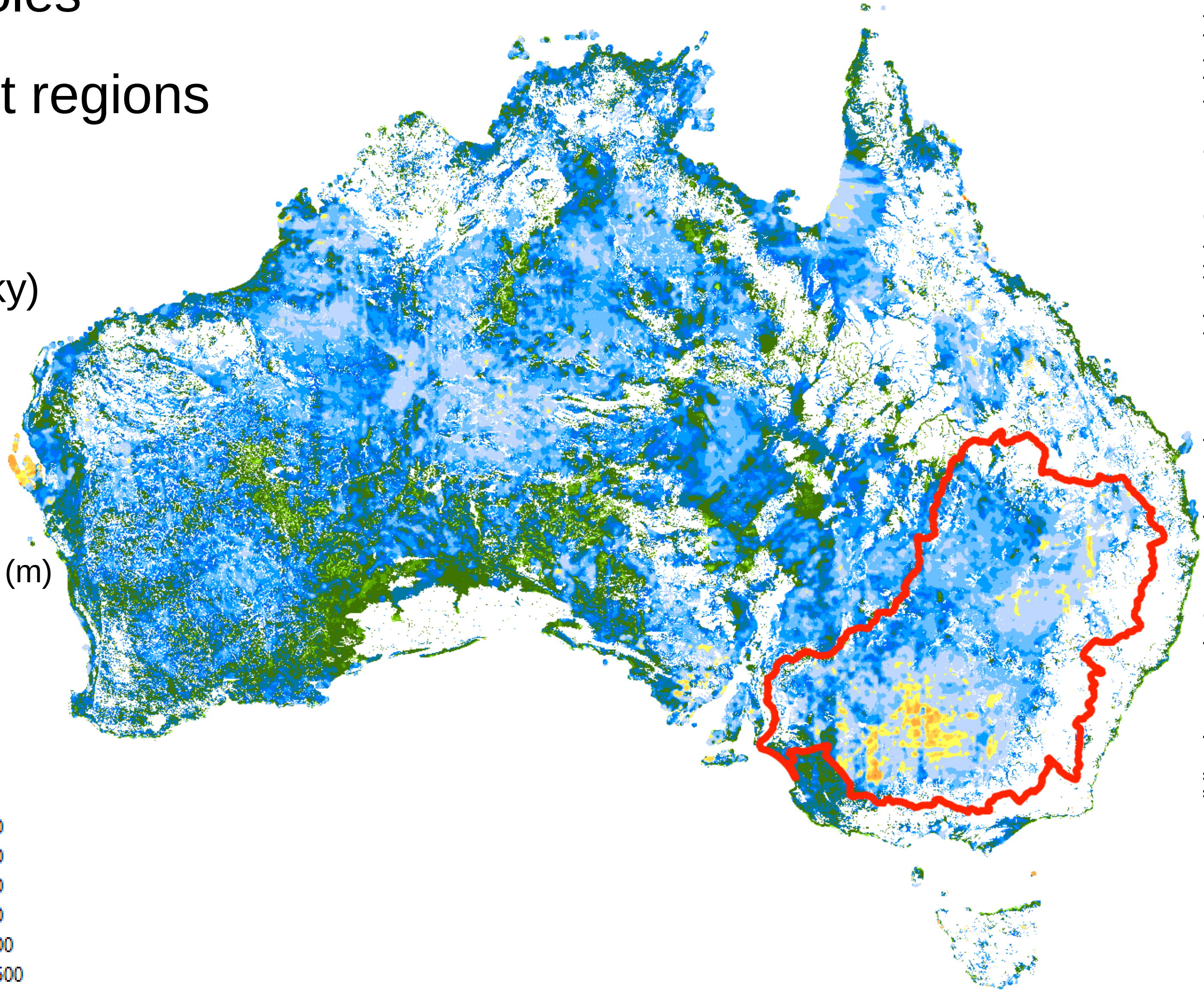


Example 3: Improving geology, focus sediment thickness

- we validated with ~40.000 boreholes
- decent results in deeper sediment regions
- example Australia

(for more information: check with Daniel Zamrsky)

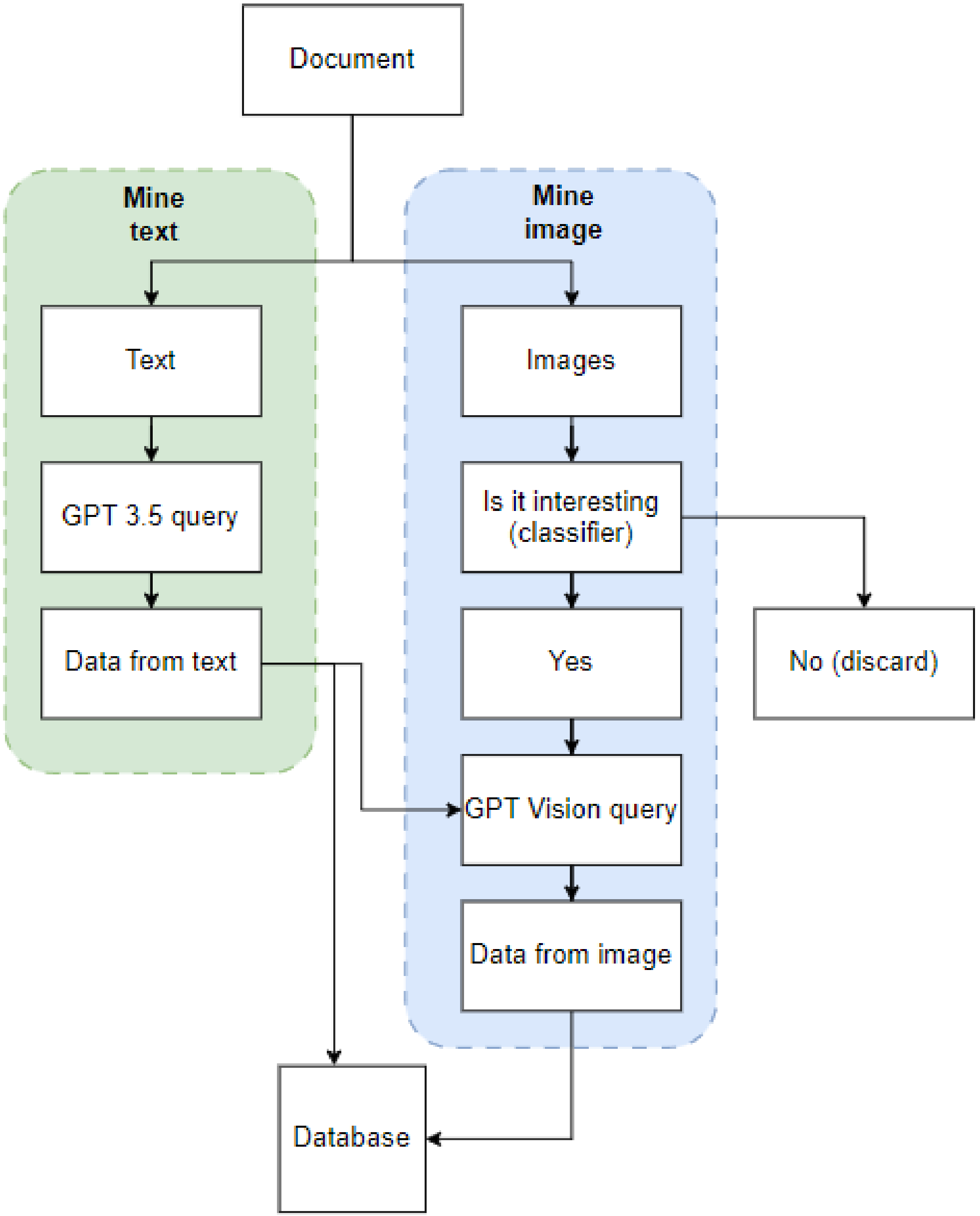




Example 4: Data mining hydrogeology

- Global models need data for **hydrogeological properties**, **validation, calibration**, etc.
- **Without enough data**, models and their **projections are highly uncertain** / people do not trust them.

- Recently available Python APIs for Large Language Models (LLMs), e.g., Open AI GPT3.5/GPT4 Vision).
- Architecture designed with FAIR practices in mind
- Short scripts, adaptable to different LLMs
- Exciting opportunities to apply these methods



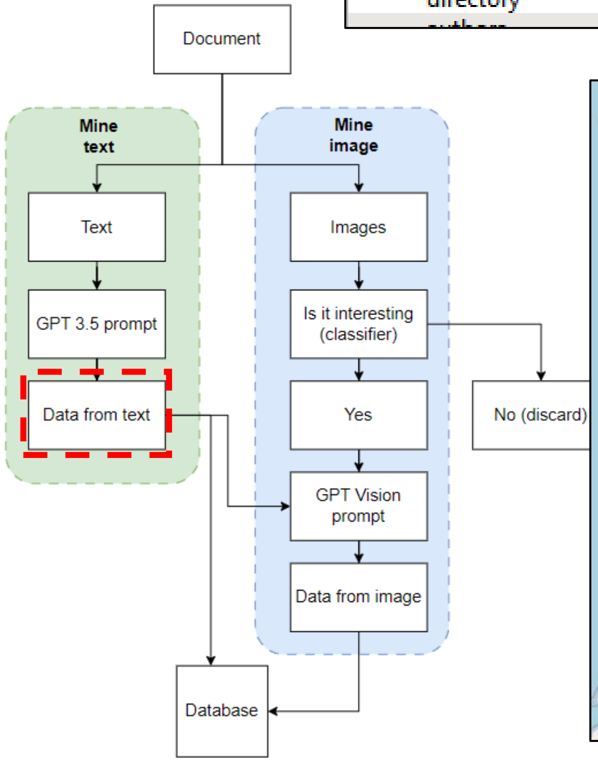
Microsoft Azure OpenAI

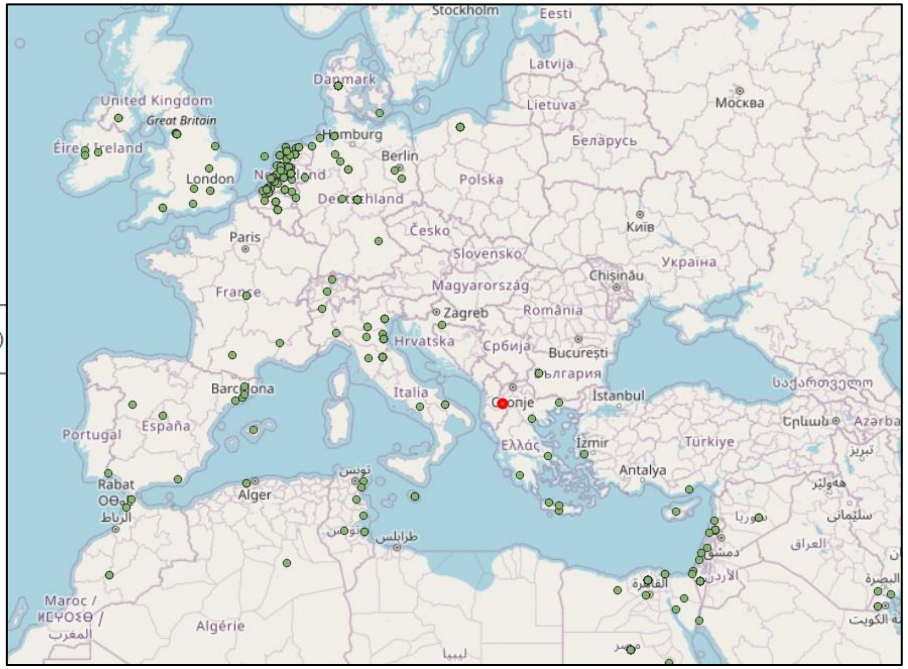
Example 4: Data mining hydrogeology

Structured data from text

- LLM response is processed using a parser and added to a structured database
- Bing maps API turns place names in coordinates
- Here: result from mining ~2000 documents
- Five orders faster & cheaper

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...

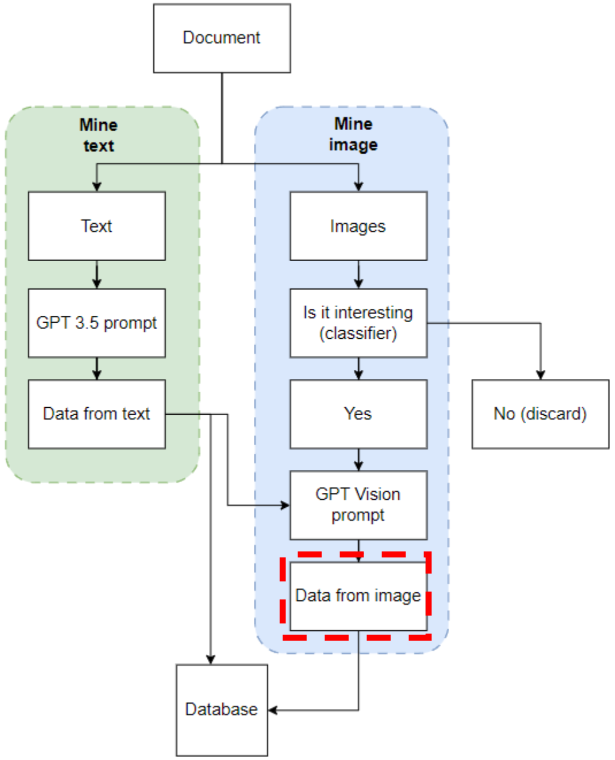




Example 4: Data mining hydrogeology

Structured data from images

- Uses multimodal capabilities (**GPT4 Vision**)
- Quantitative data needs careful structuring
- It can format and also reject incomplete or poorly structured data



Example parsed JSON output from LLM for one borehole interval

{
"top_m": 0.0,
"bot_m": 3.0,
"litho_main": "TOPSOIL",
"litho_description": "TOPSOIL GRAVEL, rare boulders",
"x_coordinate": 1661007,
"y_coordinate": 5315067,
"ground_level": 27.0,
"grid_reference": "NZTM"
},

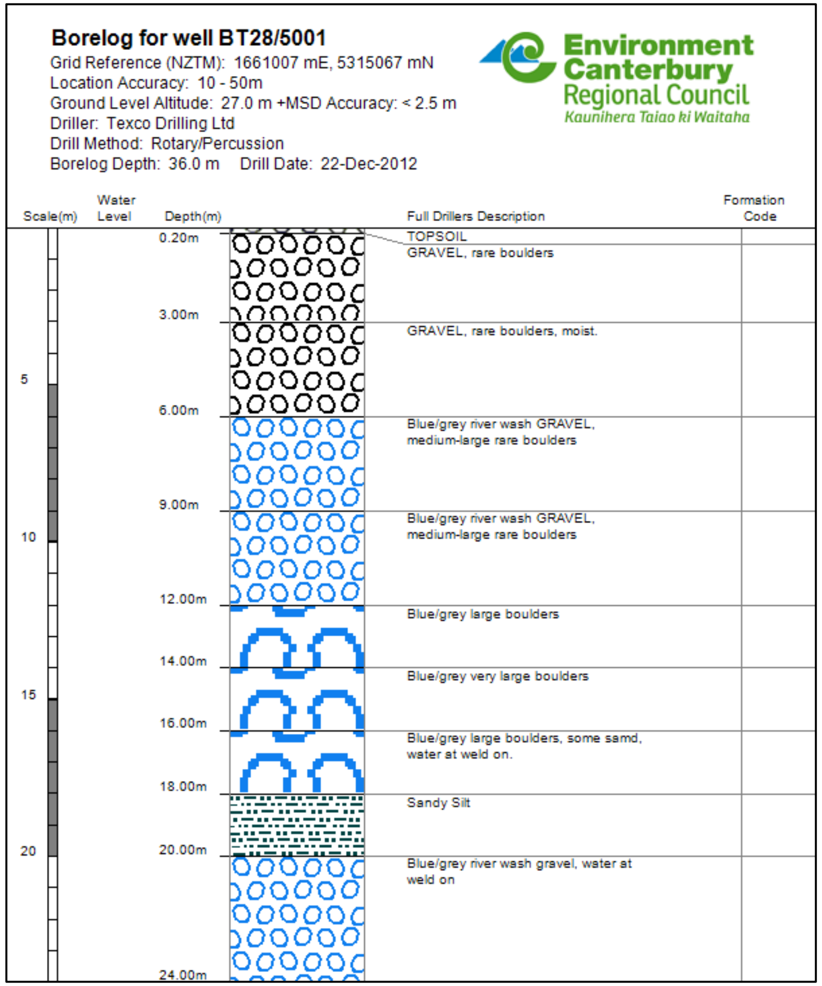
Parsed output added to database

df	top_m	bot_m	litho_main	...	ground_level	grid_reference	id
0	0.0	3.0	TOPSOIL	...	27.0	NZTM	BT28_5001.png
1	3.0	6.0	GRAVEL	...	27.0	NZTM	BT28_5001.png
2	6.0	9.0	GRAVEL	...	27.0	NZTM	BT28_5001.png
3	9.0	12.0	GRAVEL	...	27.0	NZTM	BT28_5001.png
4	12.0	15.0	BOULDERS	...	27.0	NZTM	BT28_5001.png
5	15.0	18.0	BOULDERS	...	27.0	NZTM	BT28_5001.png
6	18.0	20.0	SILT	...	27.0	NZTM	BT28_5001.png
7	20.0	24.0	GRAVEL	...	27.0	NZTM	BT28_5001.png
8	24.0	27.0	BOULDERS	...	27.0	NZTM	BT28_5001.png
9	27.0	30.0	BOULDERS	...	27.0	NZTM	BT28_5001.png
10	30.0	36.0	GRAVEL	...	27.0	NZTM	BT28_5001.png

Example 4: Data mining hydrogeology

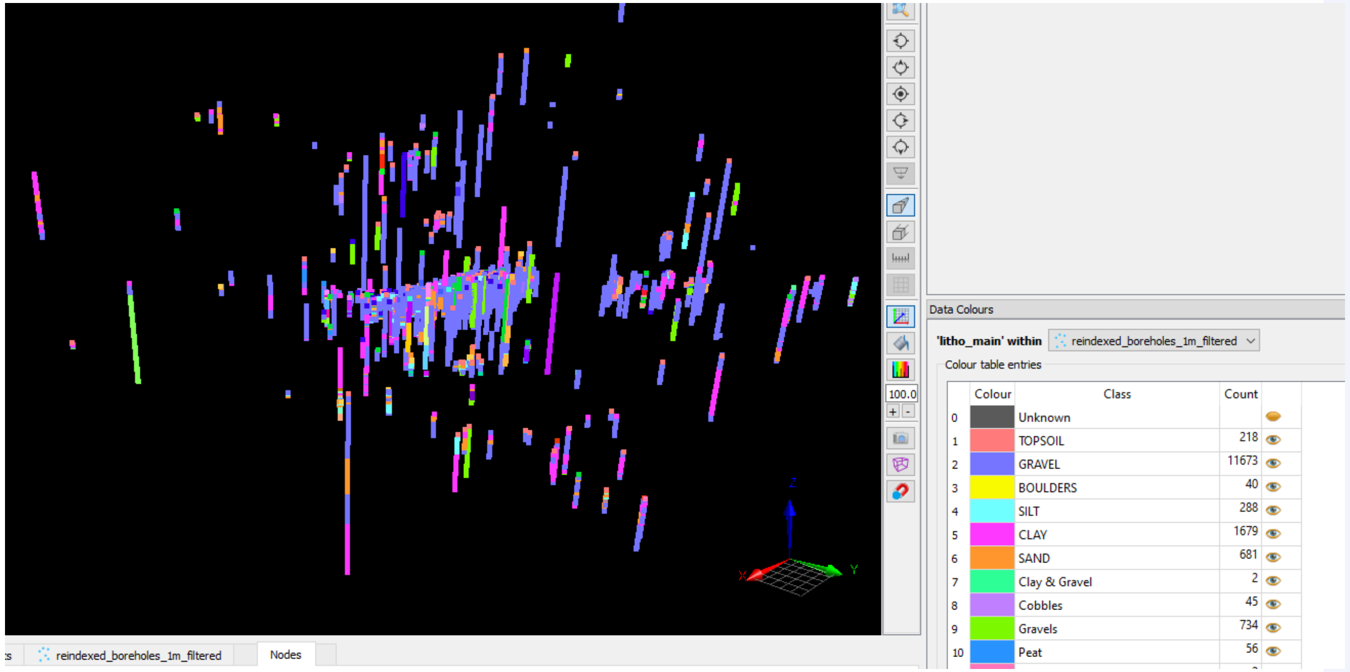
Structured data from images

Tested on ~500 borehole images from New Zealand, provided by Utrecht University


Borelog for well BT28/5001
Grid Reference (NZTM): 1661007 mE, 5315067 mN
Location Accuracy: 10 - 50m
Ground Level Altitude: 27.0 m +MSD Accuracy: < 2.5 m
Driller: Texco Drilling Ltd
Drill Method: Rotary/Percussion
Borelog Depth: 36.0 m Drill Date: 22-Dec-2012

Environment Canterbury Regional Council
Kaunihera Taiao ki Waitaha

Scale(m)	Water Level	Depth(m)	Full Drillers Description	Formation Code
		0.20m	TOPSOIL	
		3.00m	GRAVEL, rare boulders	
5		6.00m	GRAVEL, rare boulders, moist.	
		9.00m	Blue/grey river wash GRAVEL, medium-large rare boulders	
10		12.00m	Blue/grey river wash GRAVEL, medium-large rare boulders	
		14.00m	Blue/grey large boulders	
15		16.00m	Blue/grey very large boulders	
		18.00m	Blue/grey large boulders, some samd, water at weld on.	
20		20.00m	Sandy Silt	
		24.00m	Blue/grey river wash gravel, water at weld on	

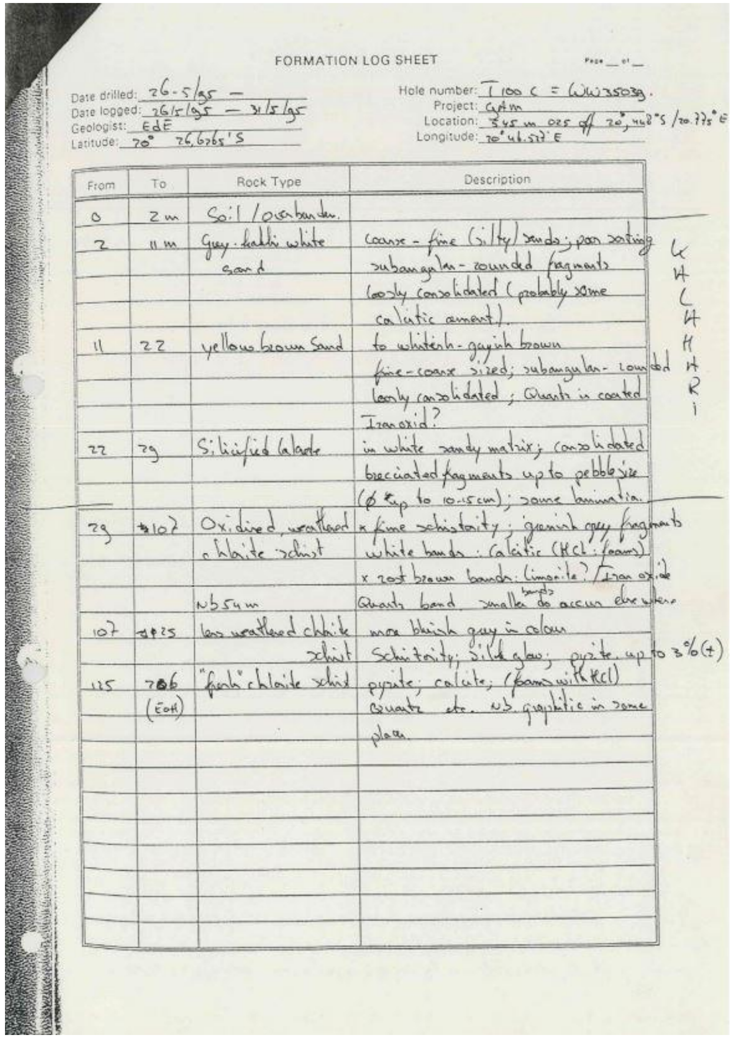


Index	top_m	bot_m	litho_main	litho_description	x_coordinate	y_coordinate	ground_level	grid_reference	id
0	0	3	TOPSOIL	TOPSOIL GRAVEL, rare boulders	1661007	5315067	27	NZTM	BT28_5001.png
1	3	6	GRAVEL	GRAVEL, rare boulders, moist.	1661007	5315067	27	NZTM	BT28_5001.png
2	6	9	GRAVEL	Blue/grey river wash GRAVEL, medium-large rare boulders	1661007	5315067	27	NZTM	BT28_5001.png
3	9	12	GRAVEL	Blue/grey river wash GRAVEL, medium-large rare boulders	1661007	5315067	27	NZTM	BT28_5001.png
4	12	15	BOULDERS	Blue/grey large boulders	1661007	5315067	27	NZTM	BT28_5001.png
5	15	18	BOULDERS	Blue/grey large boulders, some sand, water at weld on.	1661007	5315067	27	NZTM	BT28_5001.png
6	18	20	SILT	Sandy Silt	1661007	5315067	27	NZTM	BT28_5001.png
7	20	24	GRAVEL	Blue/grey river wash gravel, water at weld on	1661007	5315067	27	NZTM	BT28_5001.png
8	24	27	BOULDERS	Blue/grey large boulders and sand, water	1661007	5315067	27	NZTM	BT28_5001.png
9	27	30	BOULDERS	Blue/grey large boulders, rare white stone meal, large river wash gravel, water 1 l/s	1661007	5315067	27	NZTM	BT28_5001.png
10	30	36	GRAVEL	Blue/grey medium to large river wash gravel, water 1-2 l/s.	1661007	5315067	27	NZTM	BT28_5001.png

Example 4: Data mining hydrogeology

Structured data from images

Handles handwritten logs (mostly), also in other languages



FORMATION LOG SHEET

Date drilled: 26-5/95
Date logged: 26/5/95 – 31/5/95
Geologist: EdE
Latitude: 20° 26.626'S

Hole number: T100 C = WW3503g.
Project: GAM
Location:
Longitude:

From	To	Rock Type	Description
0	2 m	Soil / overburden.	
2	11 m	Grey-khaki white sand	coarse - fine (silty) sands; poor sorting subangular - rounded fragments loosely consolidated (probably some calcitic cement)
11	22	yellow brown Sand	to whitish-greyish brown fine-coarse sized; subangular - rounded loosely consolidated; Quartz is coated Iron oxide?
22	29	Silicified (Lahar	in white sandy matrix; consolidated brecciated fragments up to pebble size
29	102	Oxidised, weathered chlorite schist	fine schistosity; greenish grey fragments white bands: calcitic (HCl: foams)
102	125	less weathered chlorite schist	more bluish gray in colour Schistosity; pyrite, calcite
125	126 (EOH)	"fresh" chlorite schist	pyrite; calcite; (foams with HCl) Quartz etc.

top_m	bot_m	litho_main	litho_description
0	2	Soil/Organic	Coarse-fine (silty) sands, poor sorting
2	11	Grey silt with some sand	Subangular to rounded fragments, poorly consolidated (possibly some volcanic glass)
11	22	Yellow/brown Sand	Coarse to fine grain brown, ferroan coated,…
22	29	Silicified (Lahar	White sandy matrix (poorly sorted, bimodal)…
29	102	Oxidised weathered basalt	Some scoriacity, general clay fragments, wh…
102	125	Basalt	Fresh basalt, small to medium olivine phenocrysts, black glassy groundmass
125	126	Frank chlorite schist	Epidote, calcite (HCL), quartz, schistosity, silty, (possibly up to 30 (Gt))

Example 4: Data mining hydrogeology

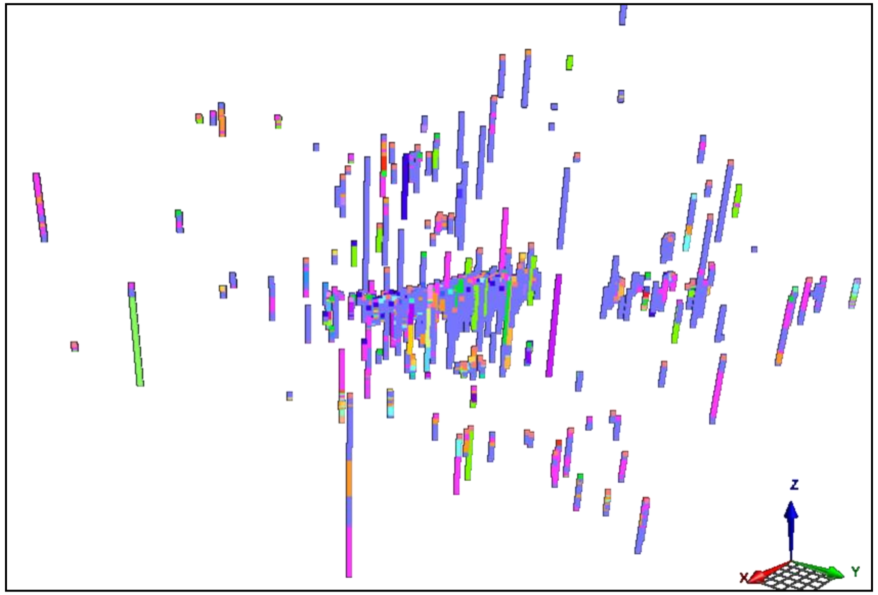
GemPy

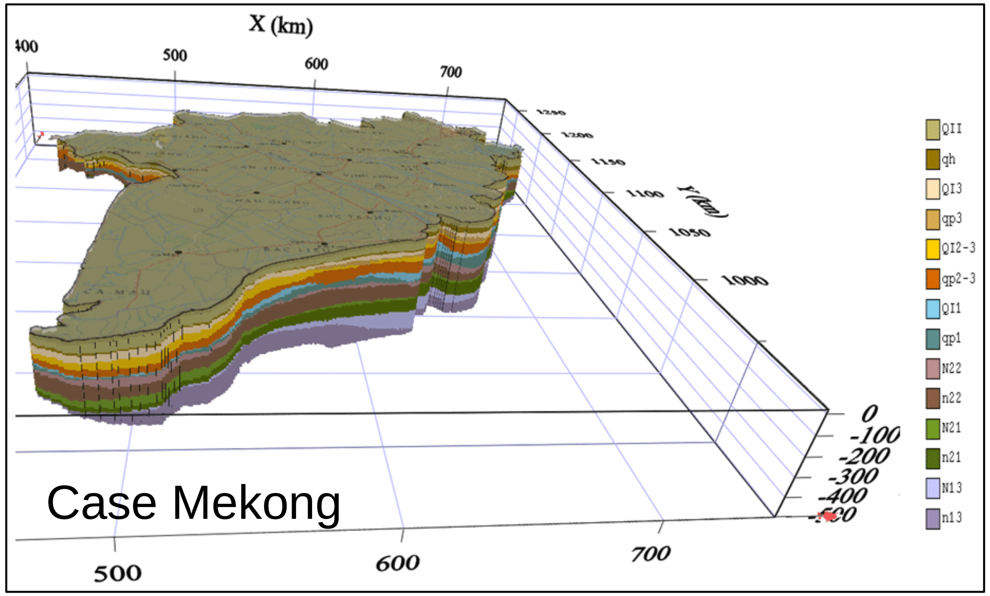
Using the data to construct large-scale hydrogeological models

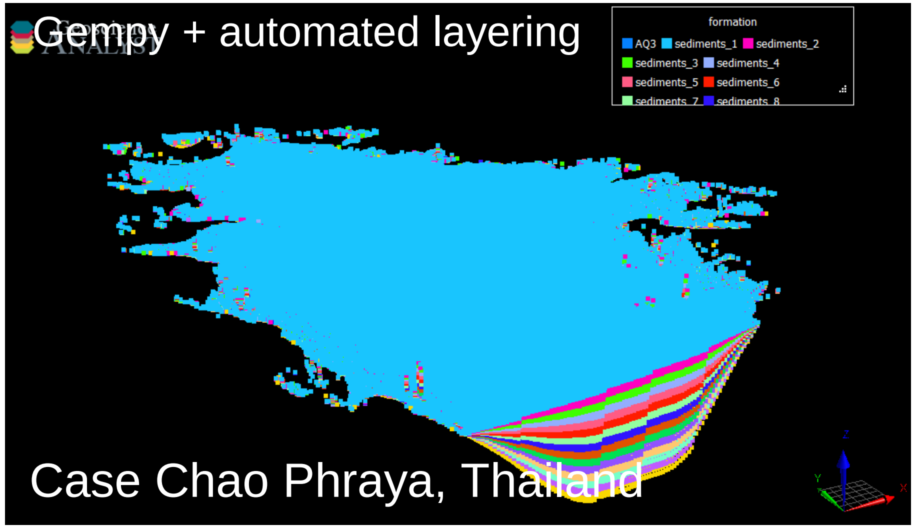
Models can (mainly) parameterised in two ways with data-mining:

1. Hydrogeological properties
2. Initial salinity distributions

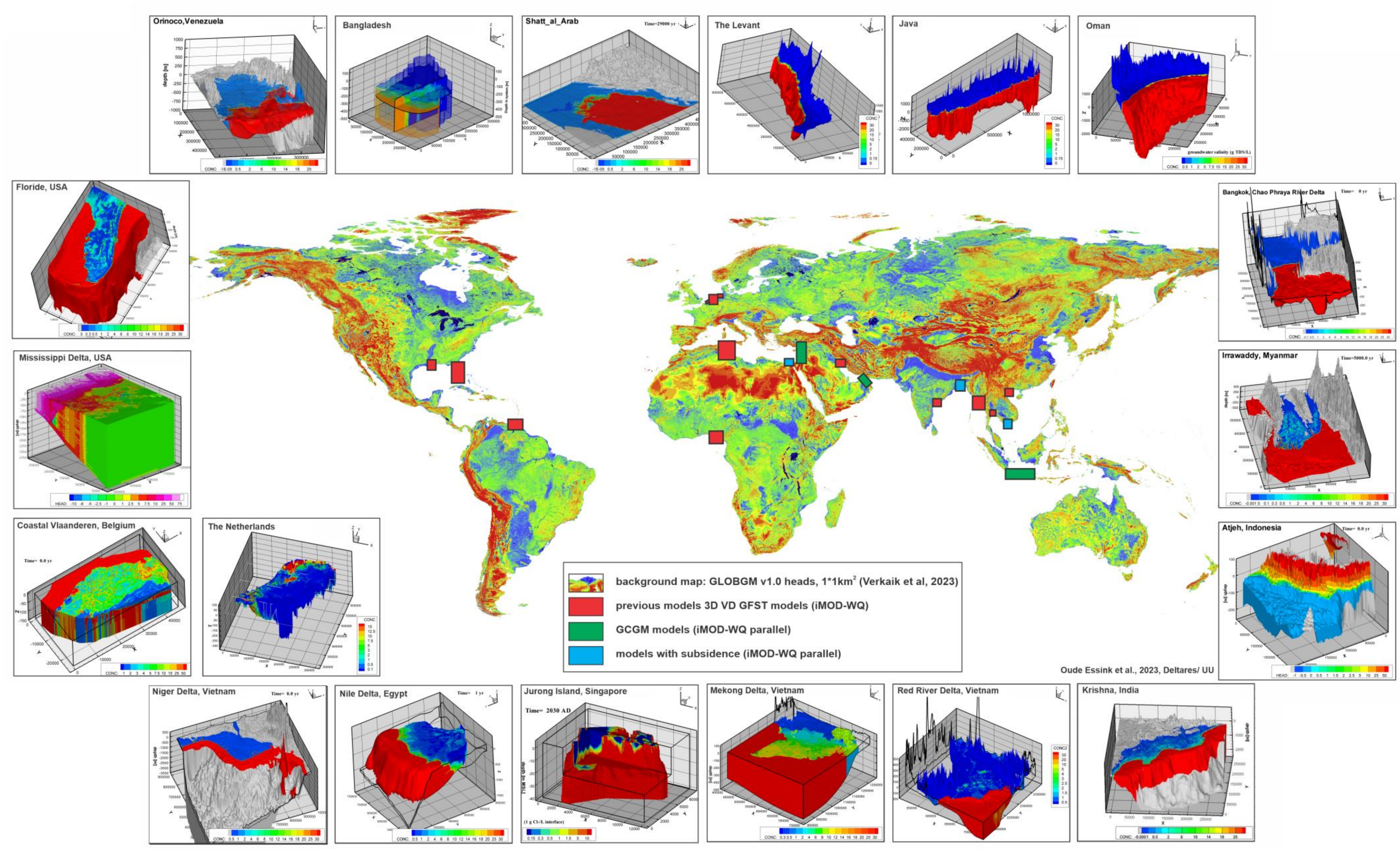
Demonstration of 3D interpolation for hydrogeological properties



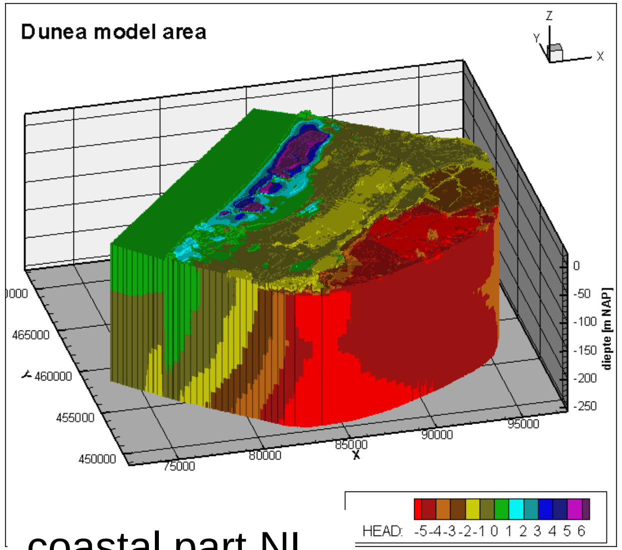




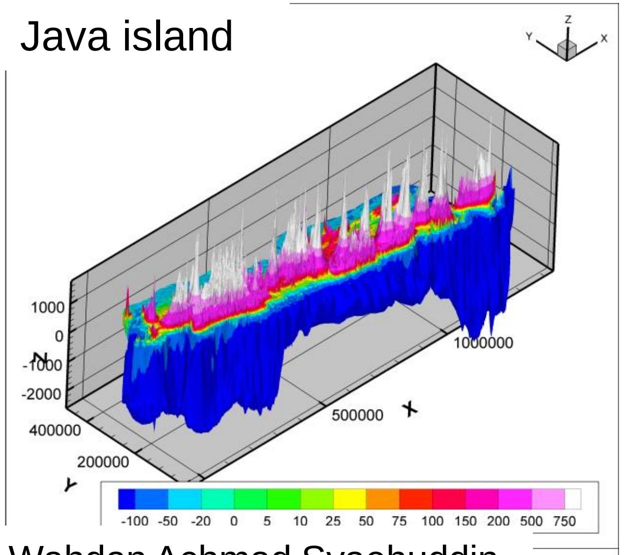
Example 5: some 1*1km^2 large-scale grondwater models



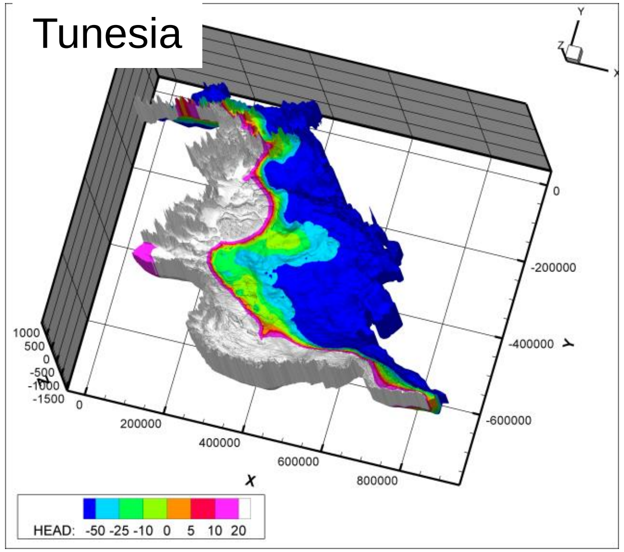
Example 5: some 1*1km^2 large-scale grondwater models

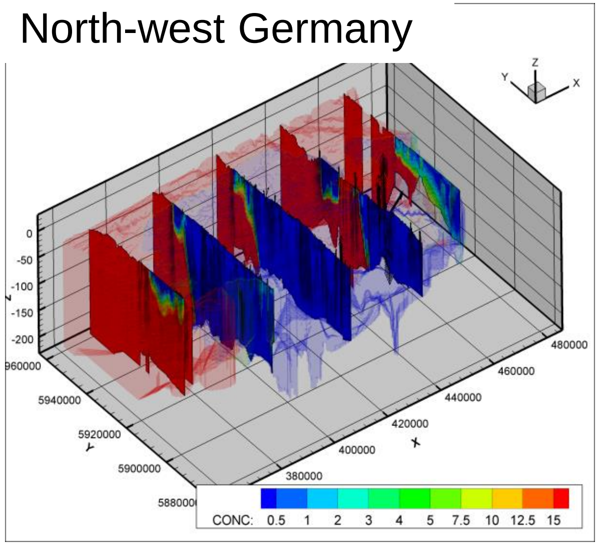


coastal part NL

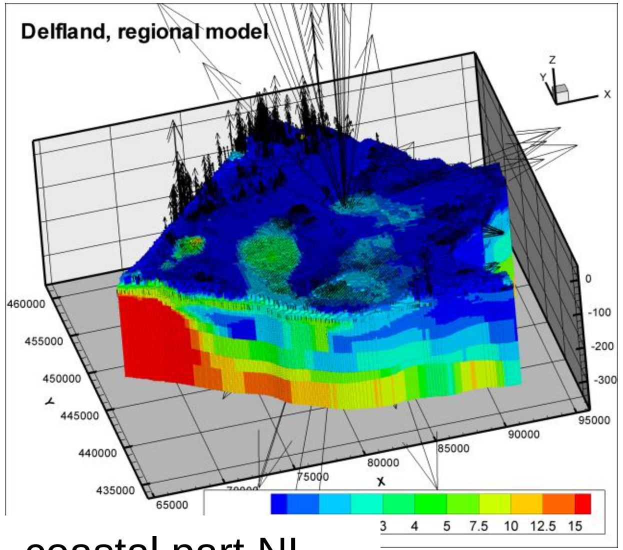


Wahdan Achmad Syaehuddin

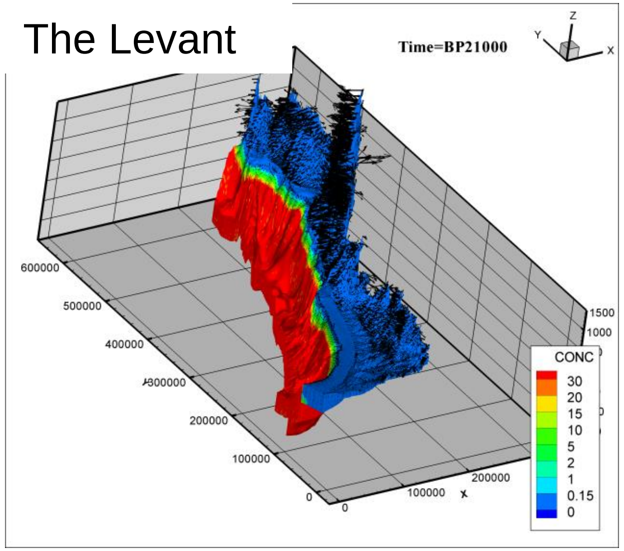




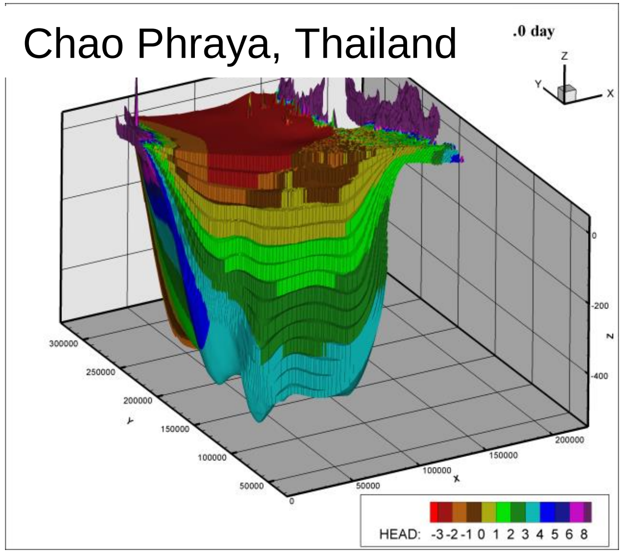
Seibert et al. WRR. 10.1029/2022WR033151

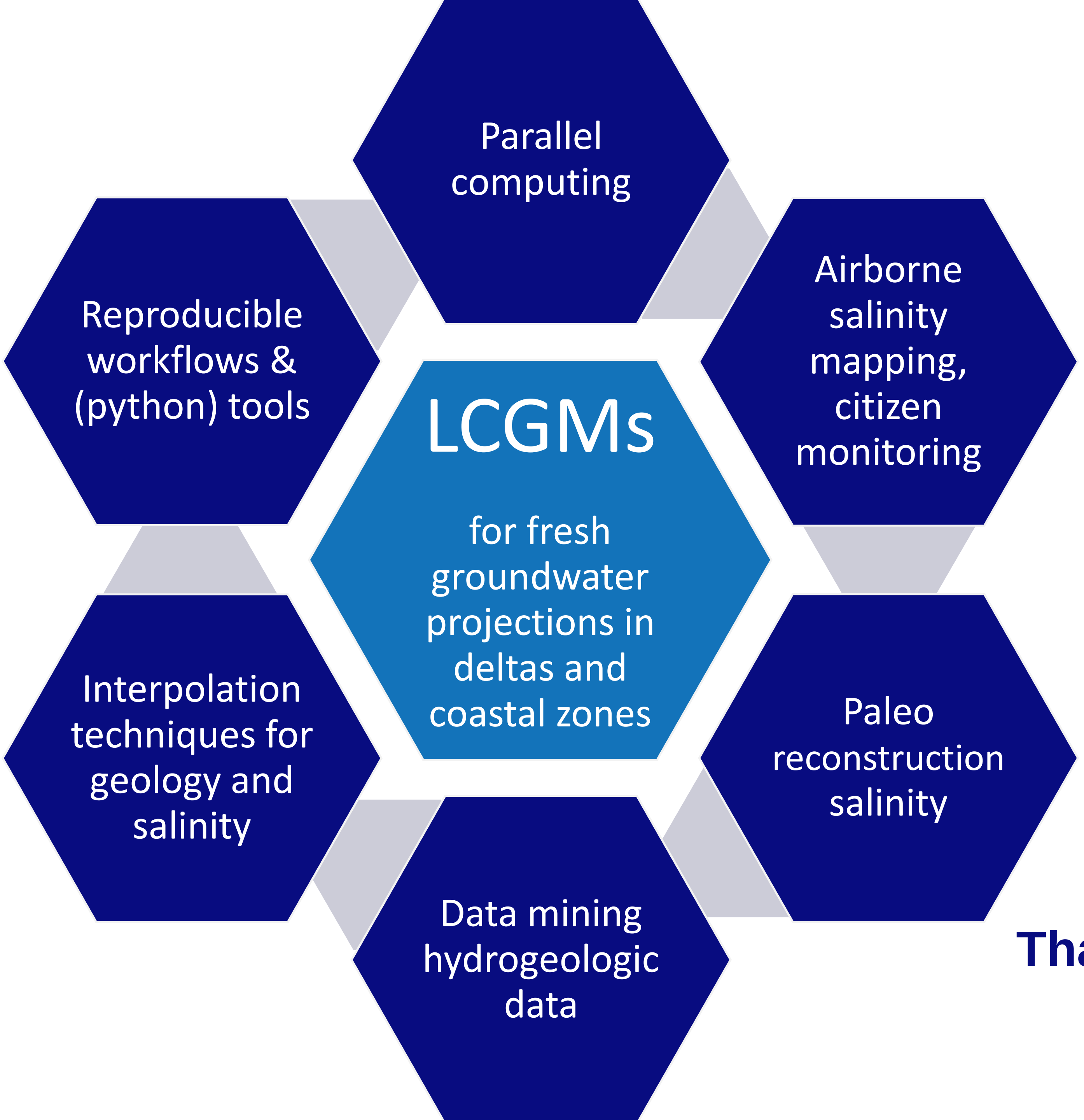


coastal part NL



Mohammed Alkurd

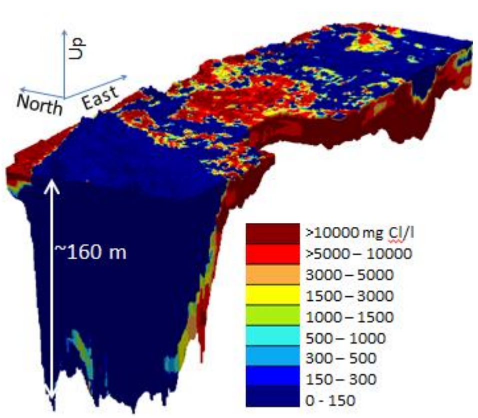




Thank for your attention

Questions?

Airborne groundwater salinity mapping



Method:
Combination helicopter measurements with knowledge about subsurface and processes in fresh-saline groundwater, and geostatistical mapping via (multiple) indicator kriging.

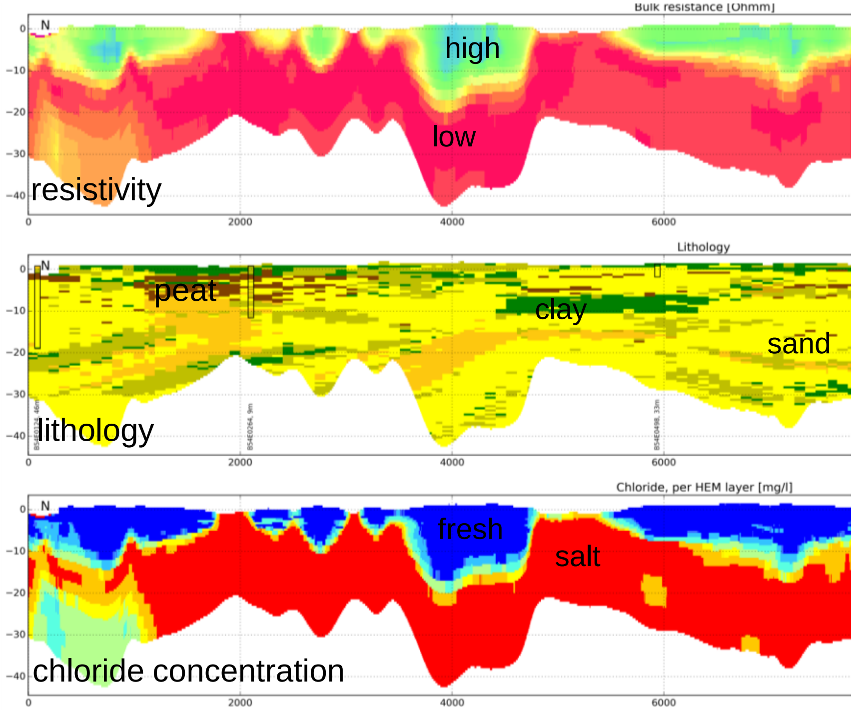
Results:

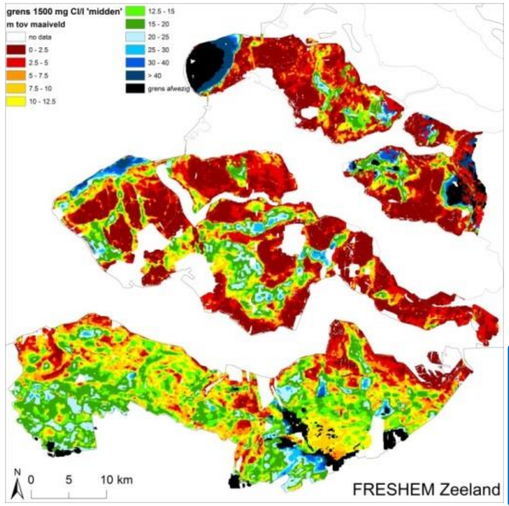
- Mapping of 3D groundwater salinity and clay layers

Applications:

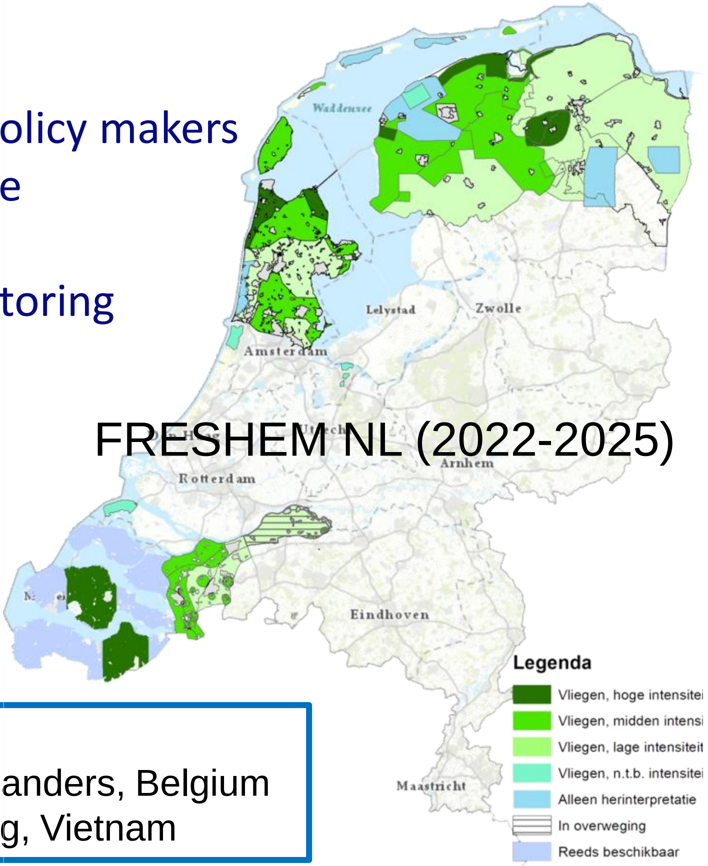
- strategic fresh groundwater users & policy makers
- support ASR (COASTAR) in coastal zone
- identify brackish water potential
- improve groundwater models & monitoring

FRESHEM Zeeland





FRESHEM NL (2022-2025)

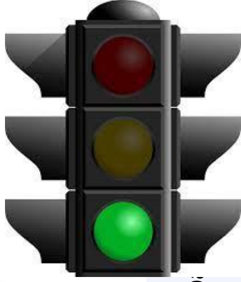


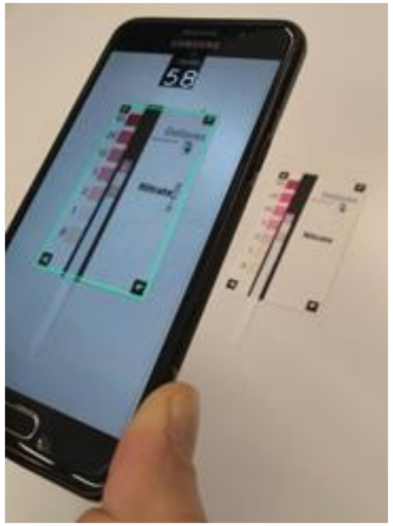
International:

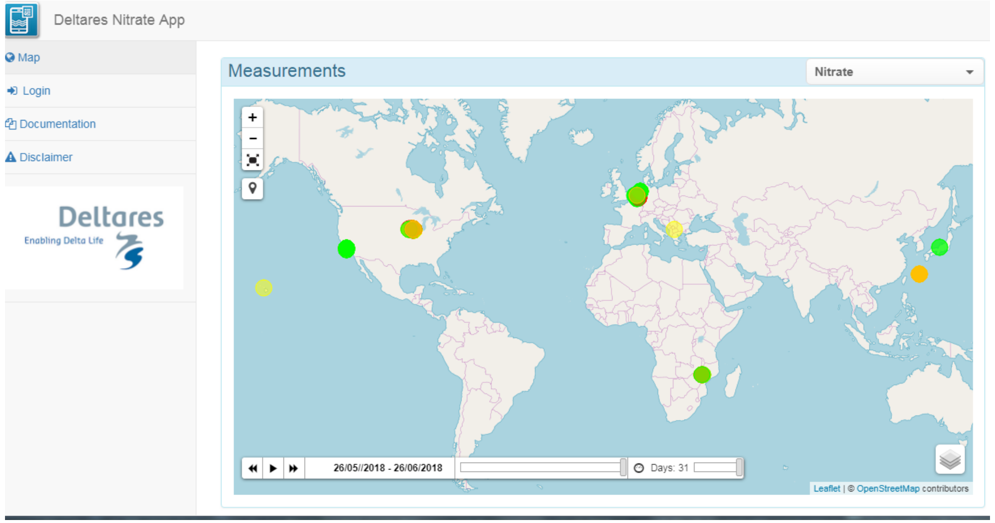
- Project in Flanders, Belgium
- Pilot Mekong, Vietnam

Building large-scale 3D coastal groundwater models with iMOD-WQ and globa

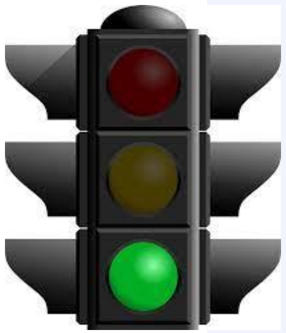
Citizen science, using simple devices and webportals



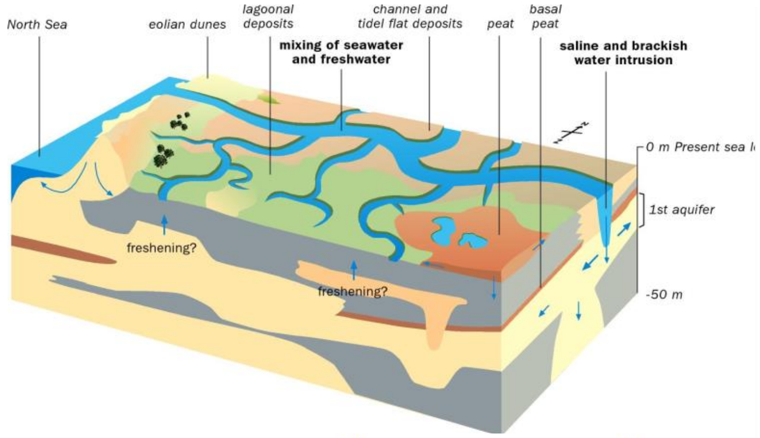


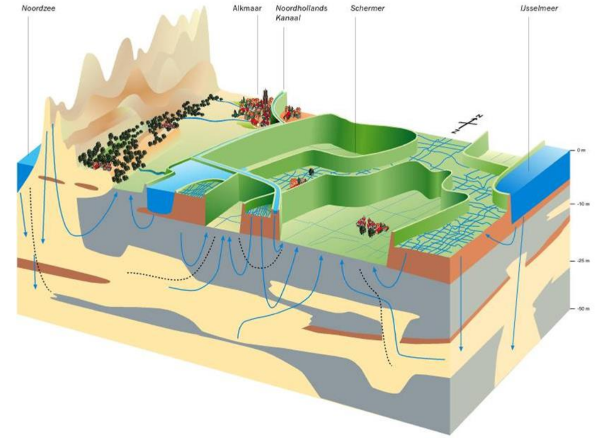


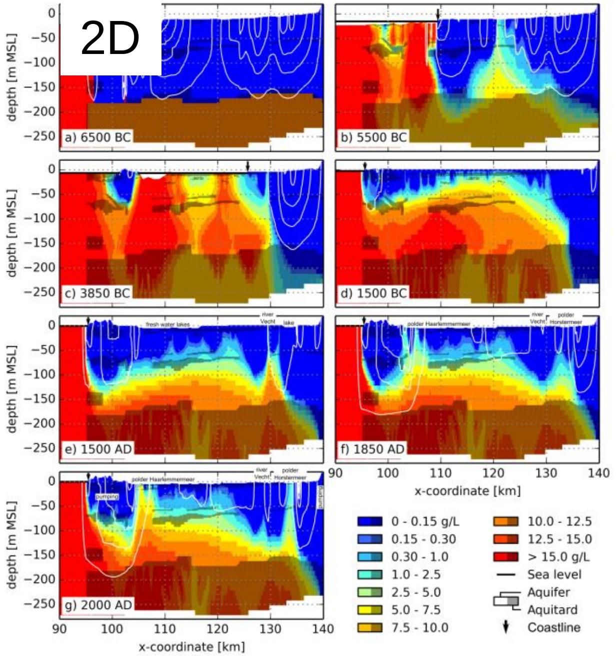
Paleo-reconstructions groundwater salinity



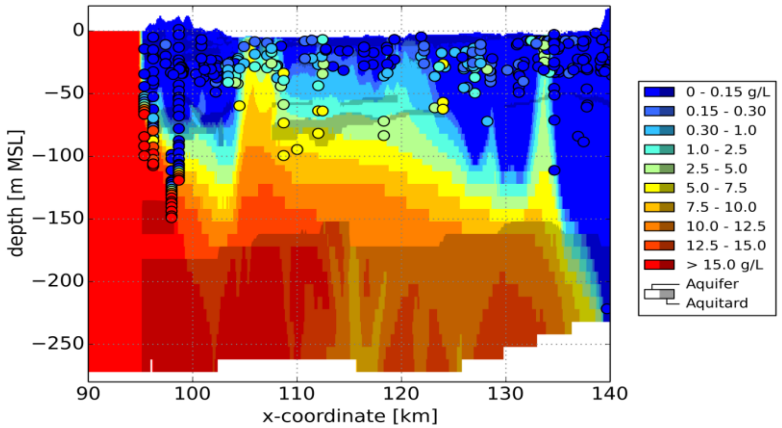
To simulate reconstructions of past hydrological conditions in (data-poor) areas, improving understanding of present groundwater salinity.







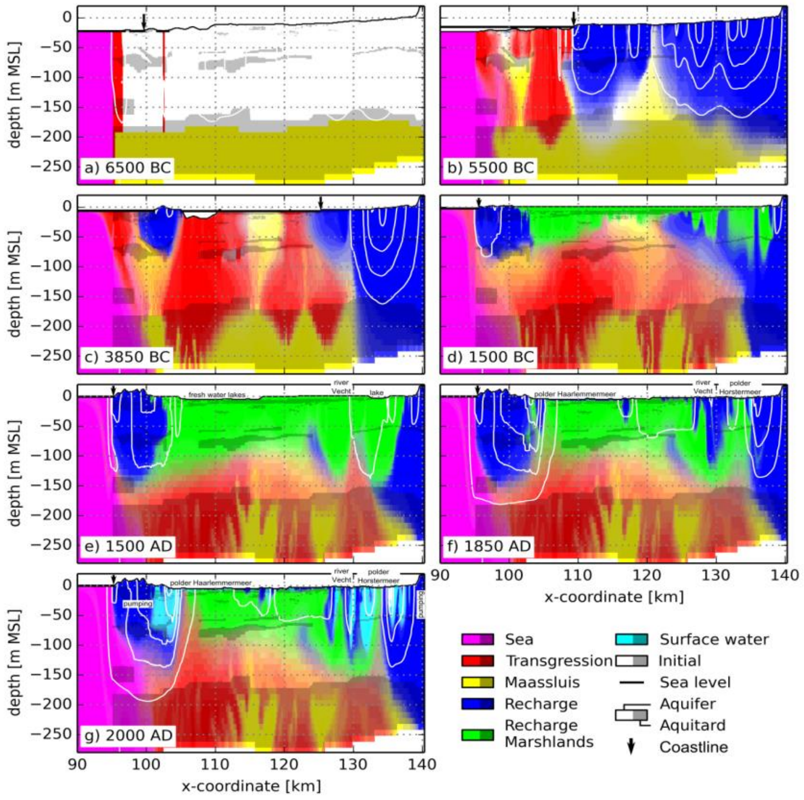
Compare with data



Delsman et al., 2014, HESS

2D → 3D

Origin of groundwater resources



Why now?

High-resolution global coastal groundwater salinity models are now possible:

1. **Parallel groundwater salinity modelling** (iMOD-WQ / SEAWAT).

2. **Fast Airborne EM groundwater salinity mapping in 3D**, (e.g., FRESHEM), citizen science data collection at high TRL.

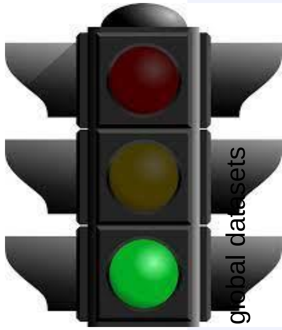
3. **Paleo reconstructions of past hydrological conditions in data-poor areas**, (possible due to parallel computer), resulting in improved understanding of present groundwater salinity.

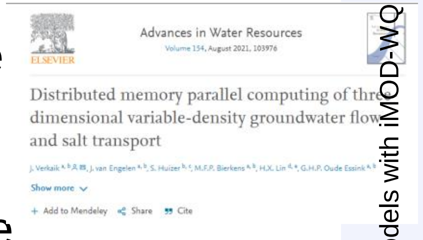
4. **More open hydrogeologic data available** (advanced **text mining**, open-source **webportals**).

5. **Advanced techniques for rapid 3D interpolation** of coastal geology and groundwater salinity, and hydrogeological model parameters.

6. **Fully scripted reproducible modelling workflows, clipping & refining** (e.g., iMOD-Python), aiding regular updating and stakeholder trust in model results.

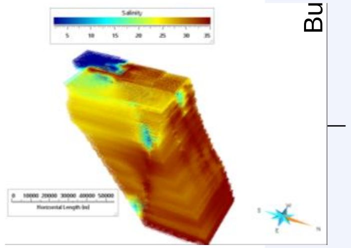
7. **And: groundwater community initiatives**, like Groundwater Model Portal (GroMoPo) (e.g. poster EGU23-12340)



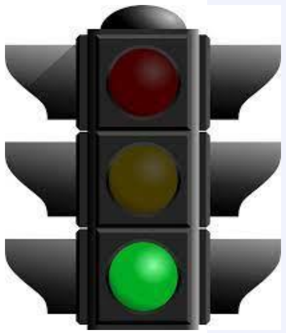

Advances in Water Resources
Volume 154, August 2021, 103976

Distributed memory parallel computing of three dimensional variable-density groundwater flow and salt transport

Show more



Paleo-reconstructions groundwater salinity

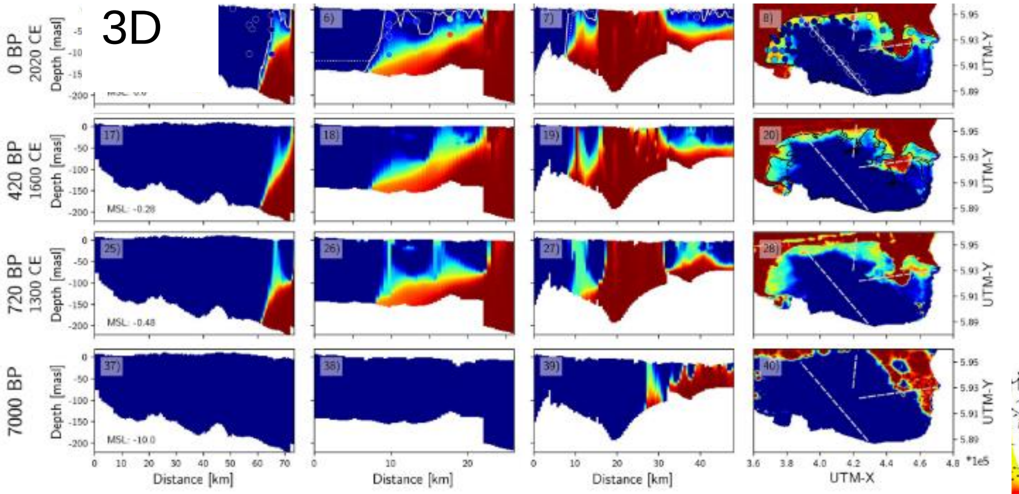


Parallel computer power is utilized to simulate 3D reconstructions of past hydrological conditions in (data-poor areas), improving understanding of present groundwater salinity.

Building large-scale 3D coastal groundwater models with iMOD-WQ and

Northwest Germany

3D

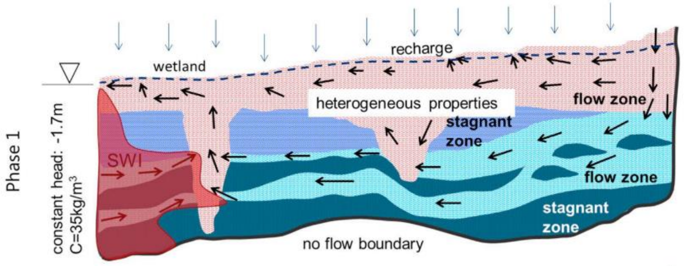


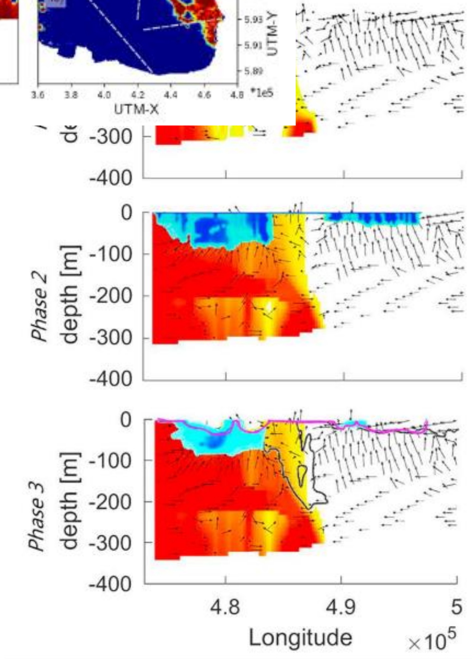
Seibert et al., 2023 WRR

Origin of sources and ~age dating

Denmark

3D

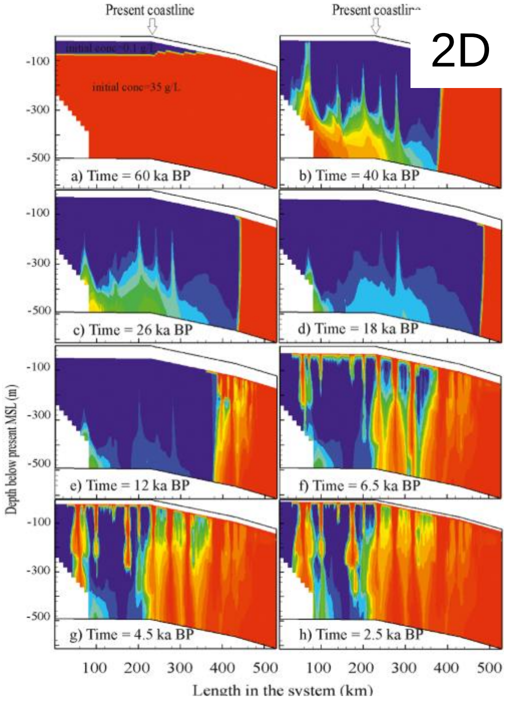




Meyer, et al., 2019

Mekong delta

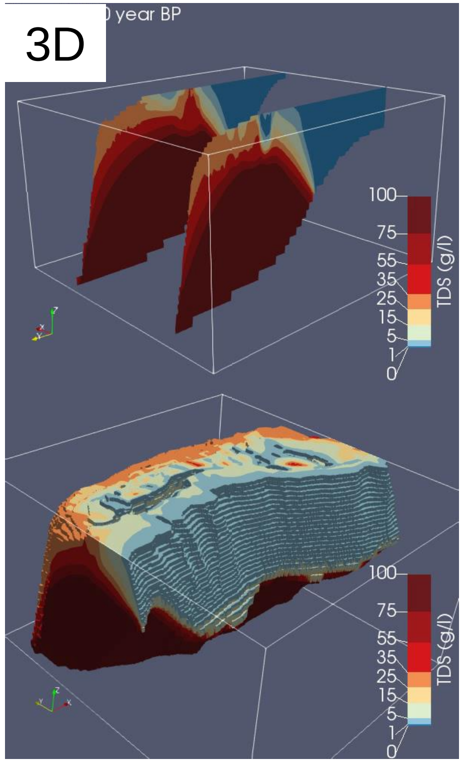
2D



Hung et al., 2019 JoH, RS

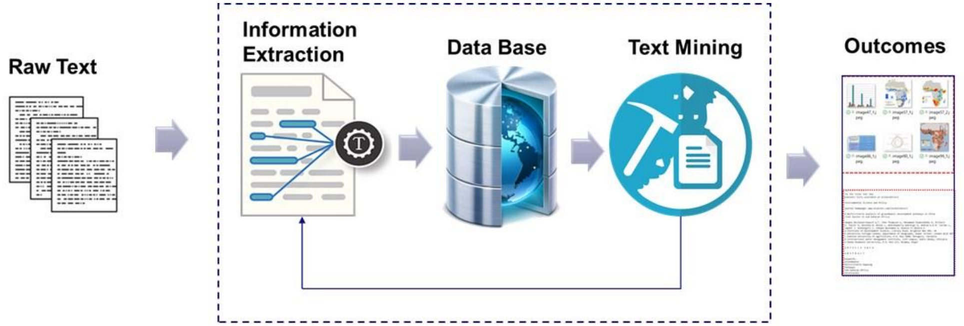
Nile delta

3D



Van Engelen et al., 2019. HESS

Data mining hydrogeology

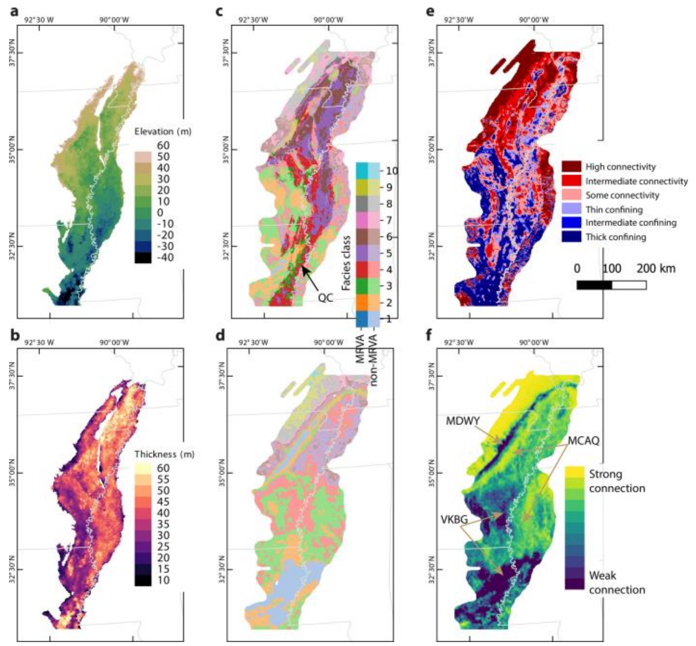


Extracting information from images.

Automated in Python using Tesseract and OpenAI API tools

PyMuPDF → Tesseract OCR → OpenAI

Input image



Output text (Tesseract)

Text editor - 21

```
35°00 N

32°30 N

92°30 W 90°00'W 92°30'W 90° 00'W 92°30'W 90° 00'W

37°30 N
37°30'N
37°30'N

35°00 N
35°00'N
35°00'N

HB High connectivity

|| Intermediate connectivity
|| Some connectivity

I) thin confining

|| Intermediate confining

|| Thick confining

32°30 N
32°30'N
32°30'N

00 200 km
```

Extracted data (OpenAI)

User: "If there are coordinates in the text, extract them as minimum and maximum coordinate pairs:

System: "Example output: X: 100, 100. Y: 100, 100"

Text editor - 21

```
Minimum coordinate pair: 32°30'N, 90°00'W
Maximum coordinate pair: 37°30'N, 92°30'W

Image: 2021_Airborne geophysical surveys of the lower Mississi_6.png
```

Data mining hydrogeology

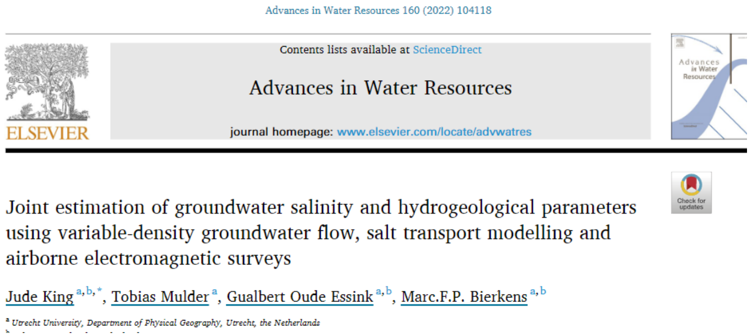
New: multi-modal GPT-4 to process multiple types of input data, combining image and textl information!

Extracting information text.

Automated in Python using OpenAI API

PyMuPDF → OpenAI

Input text



Advances in Water Resources 160 (2022) 104118

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

ELSEVIER

Joint estimation of groundwater salinity and hydrogeological parameters using variable-density groundwater flow, salt transport modelling and airborne electromagnetic surveys

Jude King[a,b,*], Tobias Mulder[a], Gualbert Oude Essink[a,b], Marc.F.P. Bierkens[a,b]

[a] Utrecht University, Department of Physical Geography, Utrecht, the Netherlands

Extracted data (OpenAI)

User: "Extract model parameters from the text in tabular format:“

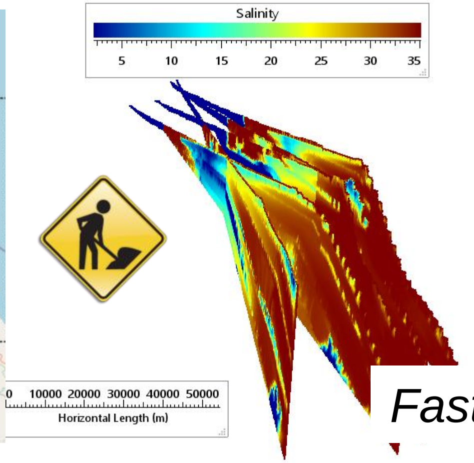
System: "Example: X,Y,Z,1,2,3"

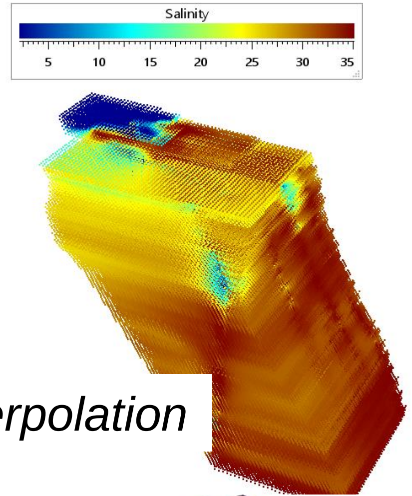
```
Initial Estimate | Actual Value | Predicted Value (difference actual)
--- | --- | --- | ---
Kh Aquifer (m/day) | 1 | 10 | 10.63291 (0.63291)
Kv Aquifer (m/day) | 0.3 | 3 | 3.312433 (0.312433)
Kh Aquitard (m/day) | 0.001 | 0.01 | 0.011838 (0.001838)
Kv Aquitard (m/day) | 0.0005 | 0.005 | 0.005831 (0.000831)
Porosity | 0.1 | 0.35 | 0.386181 (0.036181)
```

```
Kh Aquifer | Horizontal hydraulic conductivity of the aquifer | 10 (m/day)
Kh Aquitard | Horizontal hydraulic conductivity of the aquitard | 0.01 (m/day)
Kh/Kv | Anisotropy | 3.3 (aquifer) 2 (aquitard)
Porosity | Porosity | 35 (%)
Recharge winter | Recharge in winter, higher values denote ASR areas | 0.003 m/day (ASR areas), 0.0015 m/day (other areas)
Recharge summer | Recharge in summer, negative value denotes evaporation | -0.0005 m/day
Well Extraction winter | Groundwater extraction in winter | 0 m3/day per model cell
Well Extraction summer | Groundwater extraction in summer | -0.625 m3/day per model
```

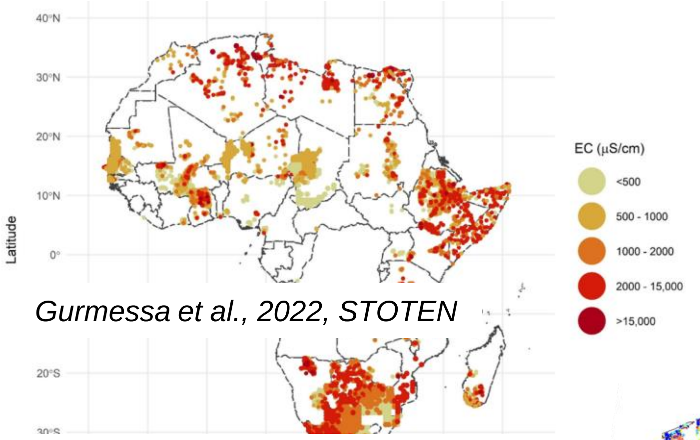
Combining techniques for 3D groundwater salinity

- Airborne EM surveys
- Text mining for pdfs and webportals
- Rapid, automated interpolations
- Paleo reconstructions modelling
- (Citizen science salinity monitoring)

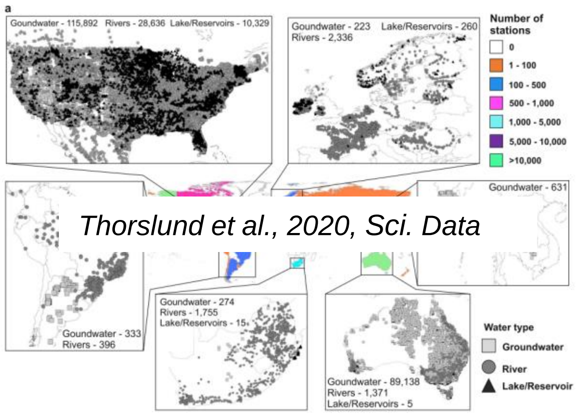




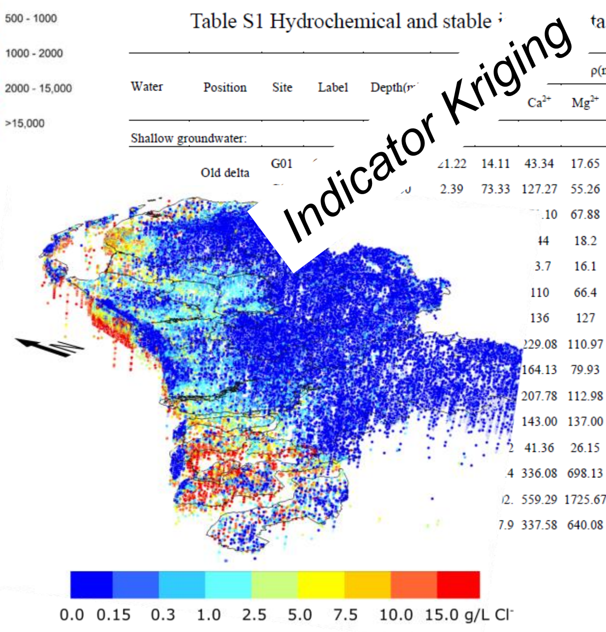
Fast interpolation



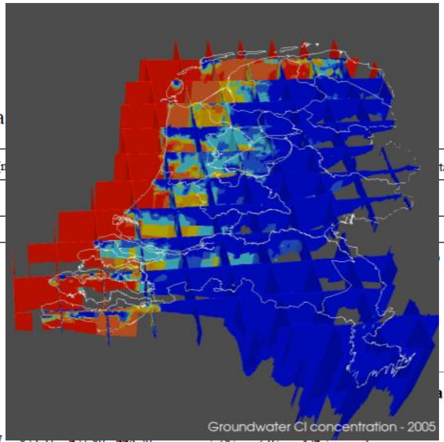
Gurmessa et al., 2022, STOTEN

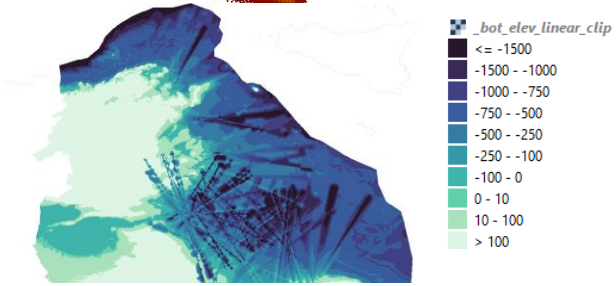


Thorslund et al., 2020, Sci. Data

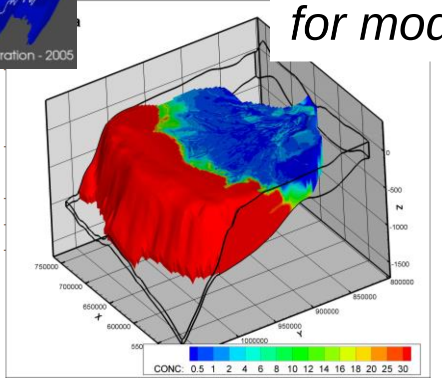


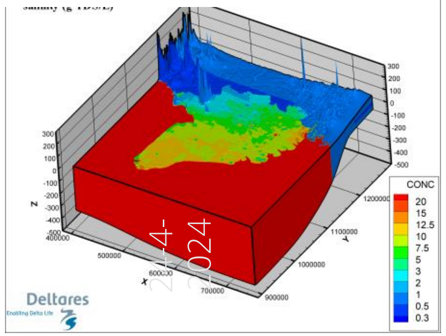
Indicator Kriging





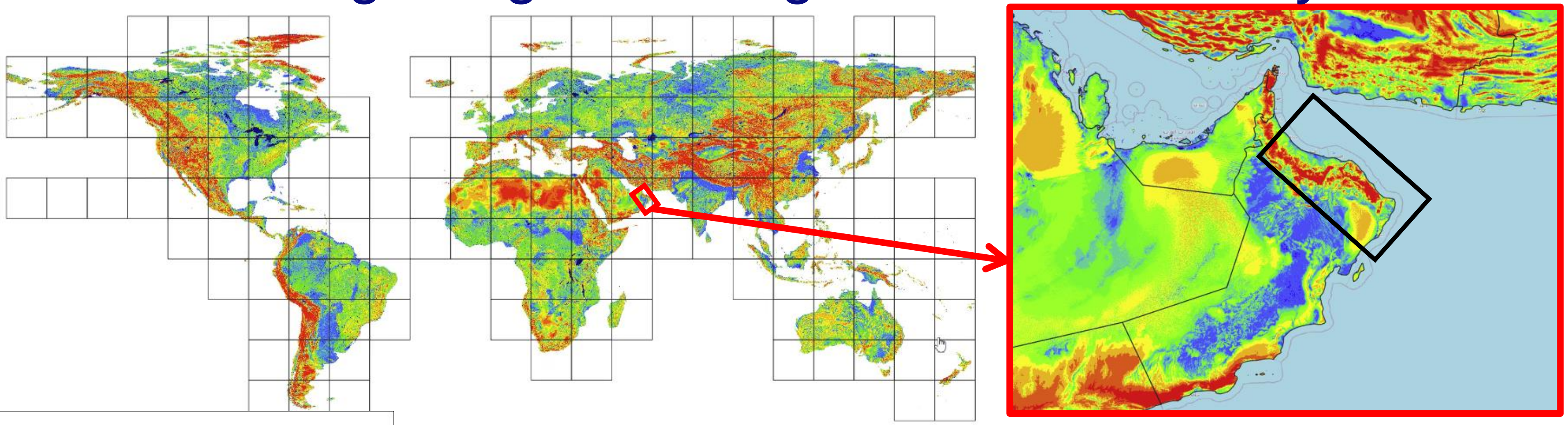
Estimate bottom elevation for modelling

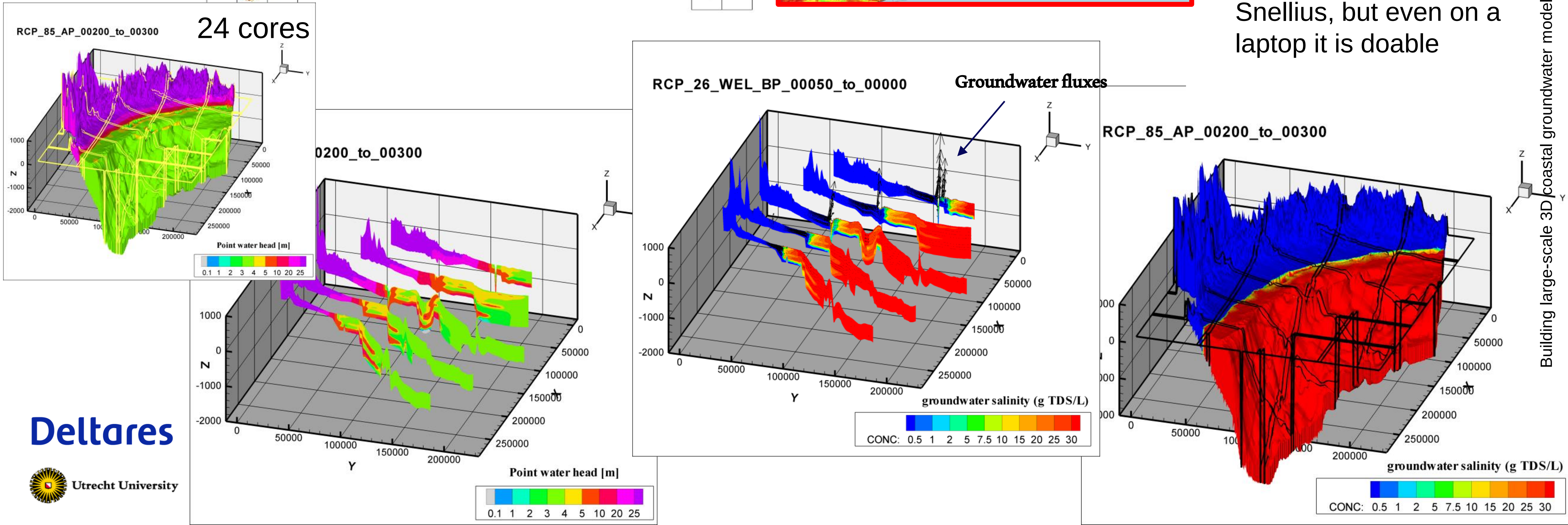




Test 1 Making a regional 3D groundwater salinity model

- Oman case
- 223*274*12 cells of 1*1km^{2};
- Simulating groundwater salinity paleo-reconstruction (120**k**yrs) and 300 yrs into the future including extractions (using PCR-GLOBWB).
- Computation time: < 1day parallel on only 24 cores; using supercomputer Snellius, but even on a laptop it is doable





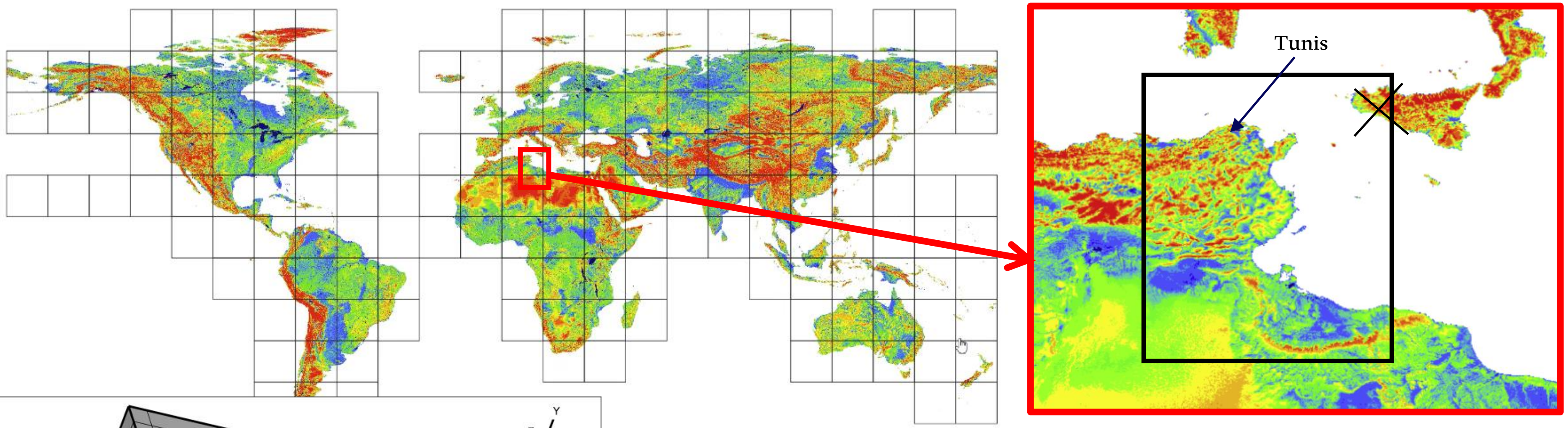
Deltares

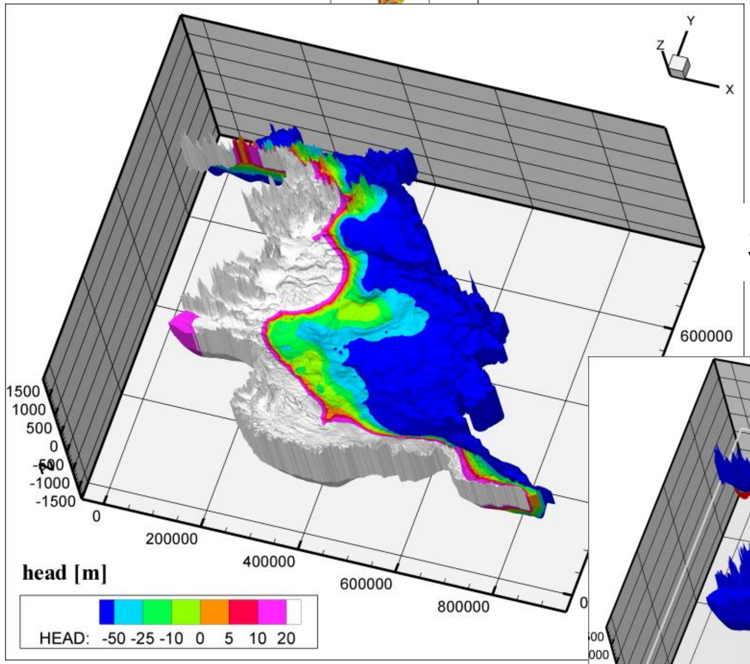
Utrecht University

Building large-scale 3D coastal groundwater models with iMOD-WQ and global datasets

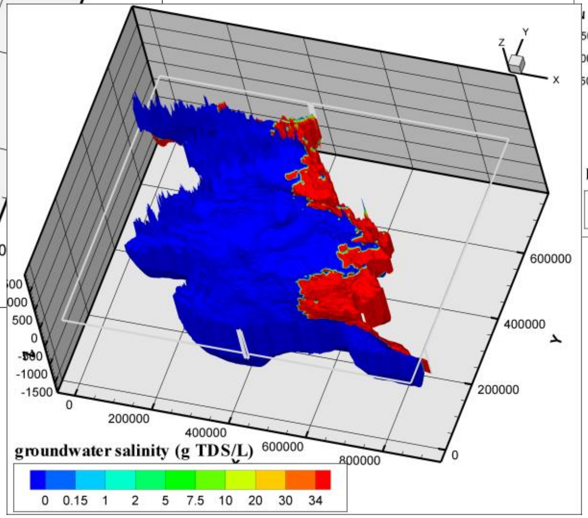
Test 2 Making a regional 3D groundwater salinity model

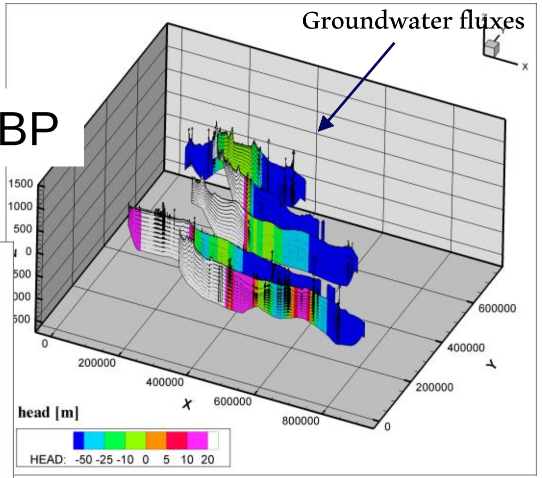
- **Tunesia-Lybia** case
- 907*747*12 cells of 1*1km^2
- Simulating groundwater salinity paleo-reconstructions. For now 30-20**k**yrs BP.
- Testing parallel on 62-128 cores; using supercomputer Snellius



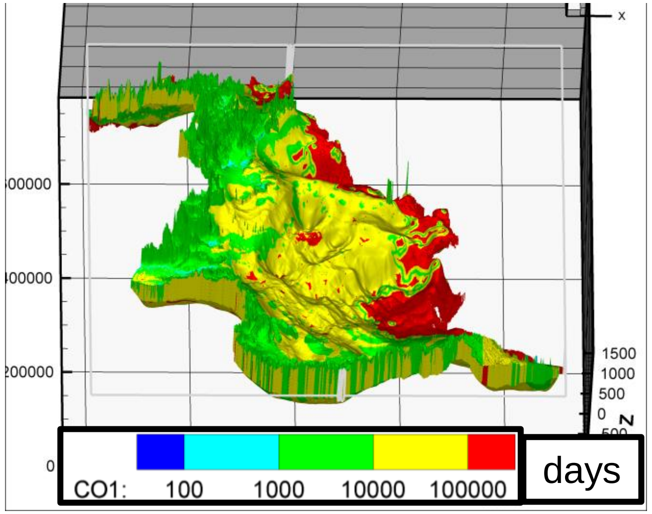


Situation 30000 yr BP





Automated STEPSIZE analysis!



- Advection,

$$\Delta t \leq \frac{R}{|v_x|/\Delta x + |v_y|/\Delta y + |v_z|/\Delta z}$$

Deltares

Utrecht University

Thank for you attention

Questions?

Gualbert Oude Essink, Daniel Zamrsky, Marc Bierkens

gualbert.oudeessink@deltares.nl

pdf on wiki freshsalt.deltares.nl

EGU23-7607 **HS8.2.6**	A study on the suitability and quantitative potential of aquifer storage and recovery and brackish water extraction in Dutch coastal areas.	**Ilja America - van den Heuvel et al**
EGU23-15557 **HS8.2.6**	Monitoring & simulation groundwater salinity due to extractions in a coastal aquifer	**Thijs Hendrikx et al**
EGU23-17249 **HS8.2.6**	Effects surface water boundary condition scaling on modelled groundwater salinity and salt fluxes	**Ignacio Farias et al**
EGU23-2859 **HS8.2.6**	Assessing impact of climate change and anthropogenic factors on future salinization; a case in Northwestern Germany)	**Stephan L. Seibert et al**
EGU23-1844 Henry Darcy Medal Lecture	**Global Water Resources and the Limits to Groundwater Use**	**Marc Bierkens**

And PICO Zamrskly et al.: **EGU23-11444** HS8.2.6

More information:

Parallel SEAWAT, imod-python and 3D viewer:

- https://oss.deltares.nl/web/imod/about-imod5
  - Verkaik, J. et al., 2021. Distributed memory parallel computing of three-dimensional variable-density groundwater flow and salt transport. Adv. Water Resour. 154, 103976. https://doi.org/10.1016/j.advwatres.2021.103976
  - https://deltares.github.io/iMOD-Documentation/python_index.html iMOD Python
  - https://deltares.github.io/iMOD-Documentation/viewer_index.html iMOD Viewer

Reproducibility and transparency, Gitlab

  - https://gitlab.com/deltares/imod/nhi-fresh-salt
  - Delsman, J.R. et al 2023. Reproducible construction of a high-resolution national variable-density groundwater salinity model for the Netherlands. Environ. Model. Softw. 105683. https://doi.org/10.1016/j.envsoft.2023.105683

- 3D Paleo-reconstruction groundwater salinity and iMOD-WQ
  - Seibert, S.L. et al., 2023. Paleo-hydrogeological modeling to understand present-day groundwater salinities in a low-lying coastal groundwater system (Northwestern Germany). Water Resour. Res. https://doi.org/https://doi.org/10.1029/2022WR033151
  - Van Engelen, J., Verkaik, J., King, J., Nofal, E.R., Bierkens, M.F.P., Oude Essink, G.H.P., 2019. A three-dimensional palaeohydrogeological reconstruction of the groundwater salinity distribution in the Nile Delta Aquifer. Hydrol. Earth Syst. Sci. 23, 5175–5198. https://doi.org/10.5194/hess-2019-151

Utrecht University

Deltares

Orange issues

- Calibration, validation, verification.
- Tekst mining: IPR of articles.
- Interferences with local hydrogeological communities, some same regional scale.

Developments into LCGM version 1.0