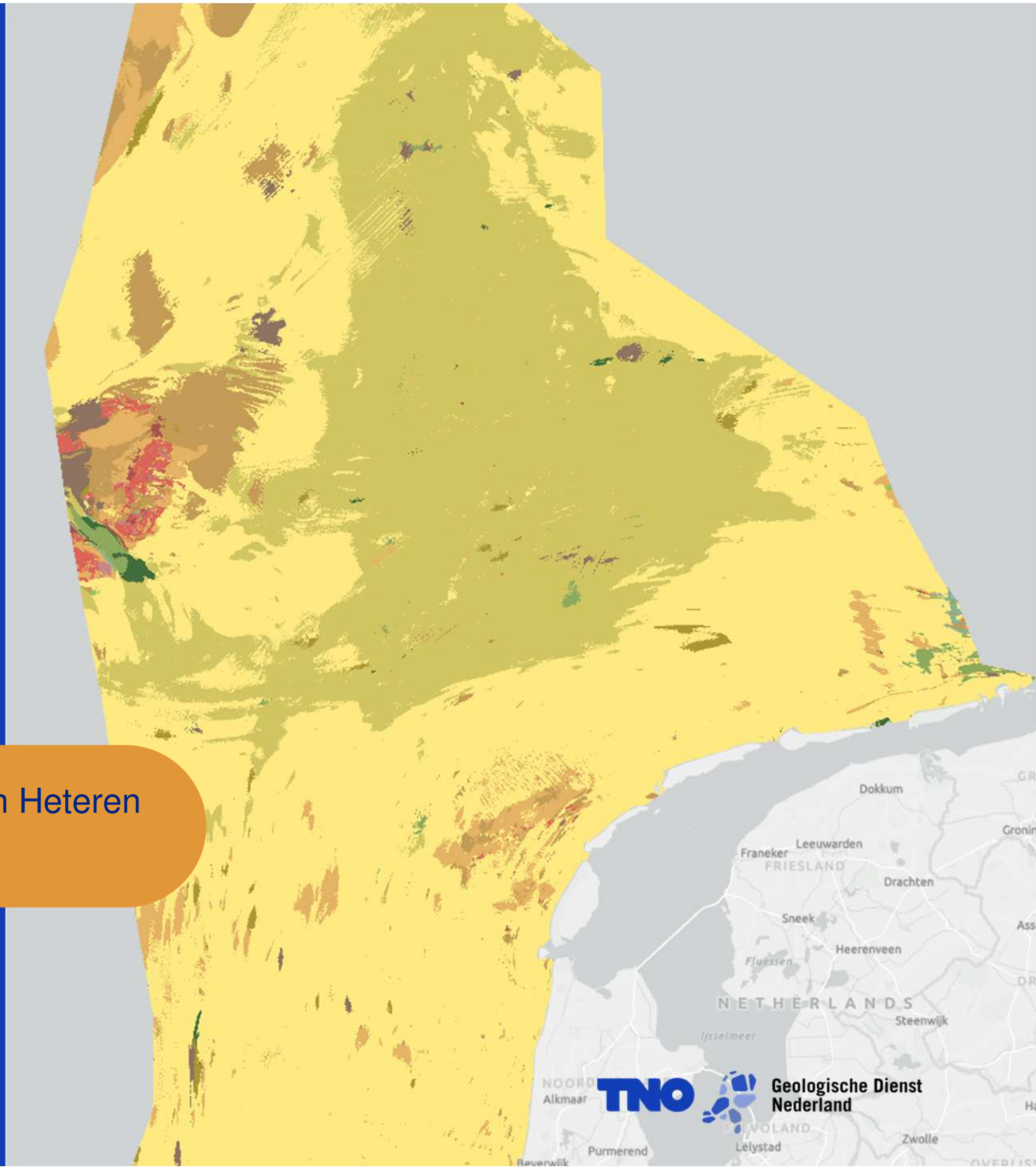
Mapping seabed sediments in the Dutch North Sea with AI

Willem Dabekaussen, Marcel Bakker, Jelte Stam, Sytze van Heteren
TNO – Geological Survey of the Netherlands

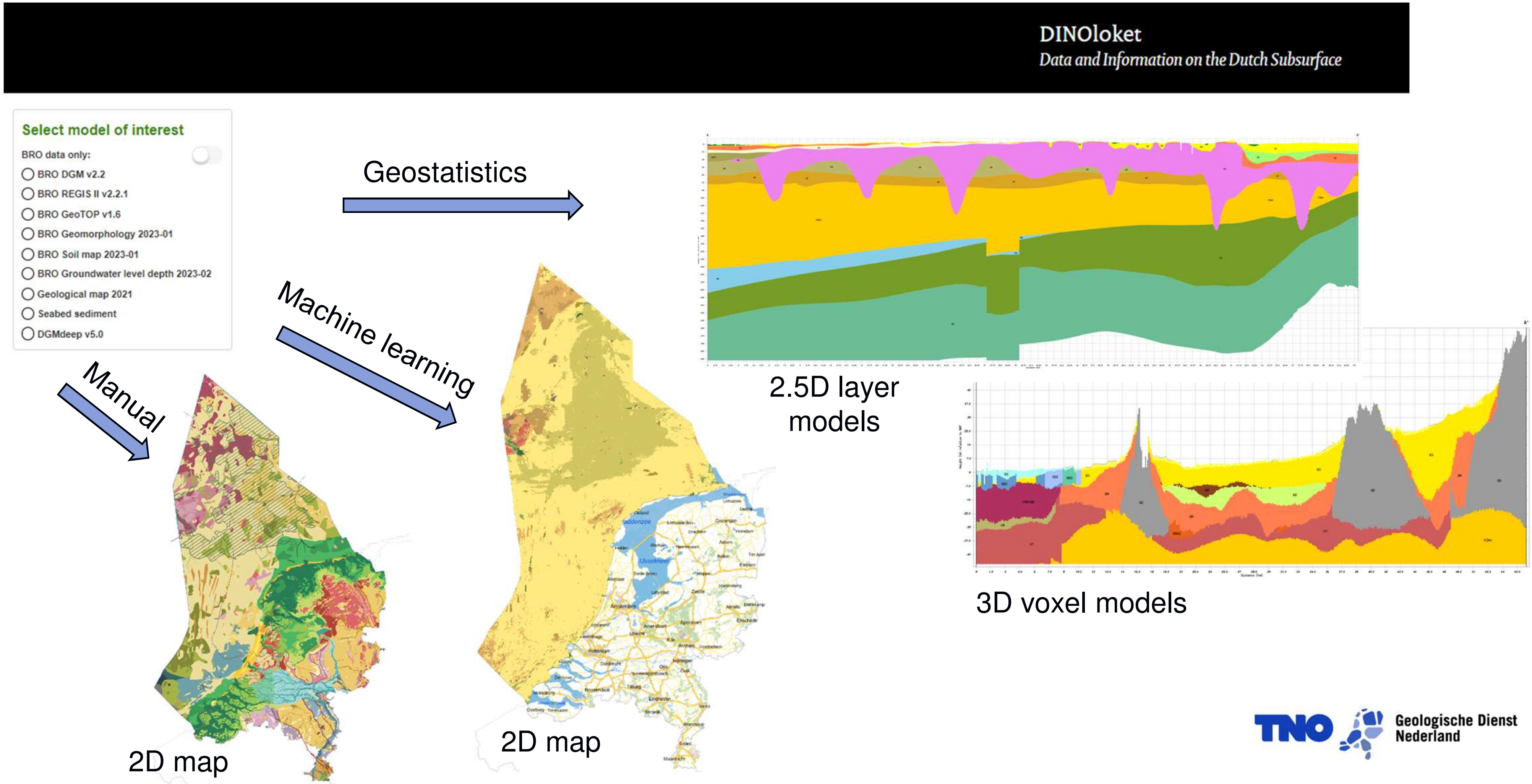
7th European meeting on 3D geological modelling - Warsaw



Geologische Dienst Nederland

TNO-GDN subsurface models

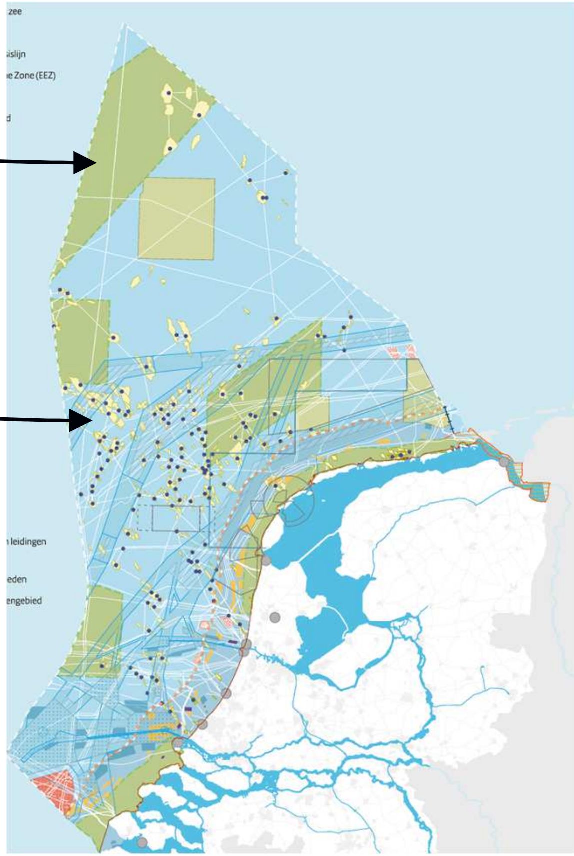
www.dinoloket.nl



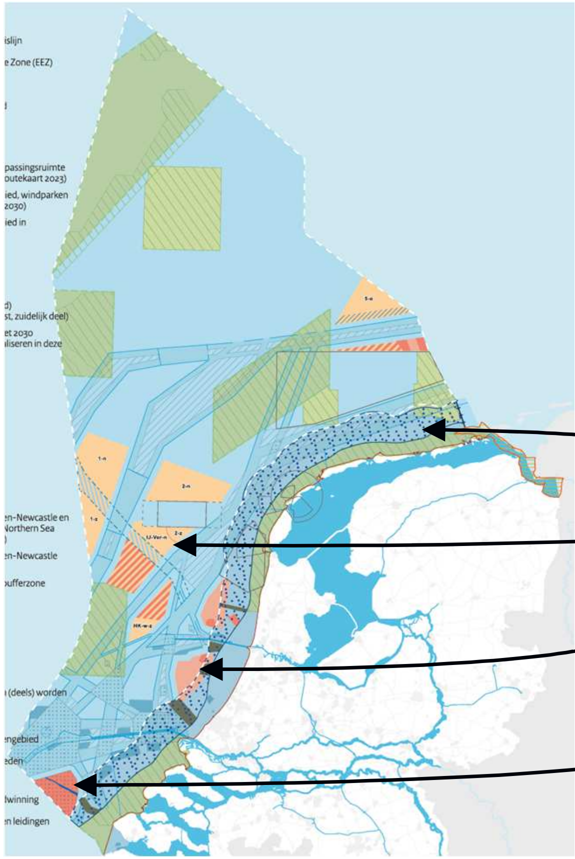
TNO Geologische Dienst Nederland

Transforming North Sea infrastructure

Current use

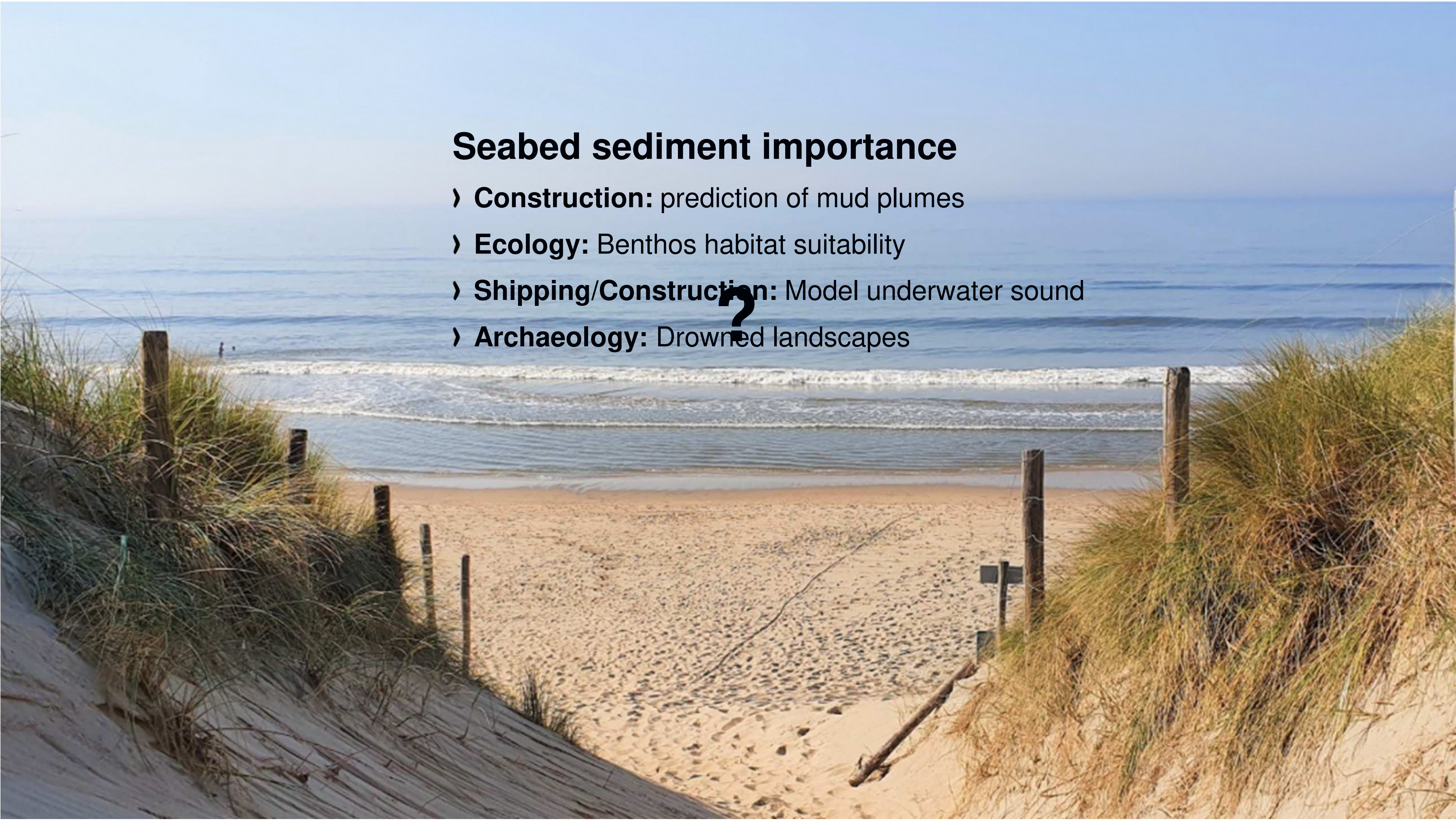


Near future



www.noordzeeloket.nl

Geologische Dienst Nederland



Seabed sediment importance

- **Construction:** prediction of mud plumes
- **Ecology:** Benthos habitat suitability
- **Shipping**/**Construction:** Model underwater sound
- **Archaeology:** Drowned landscapes

?

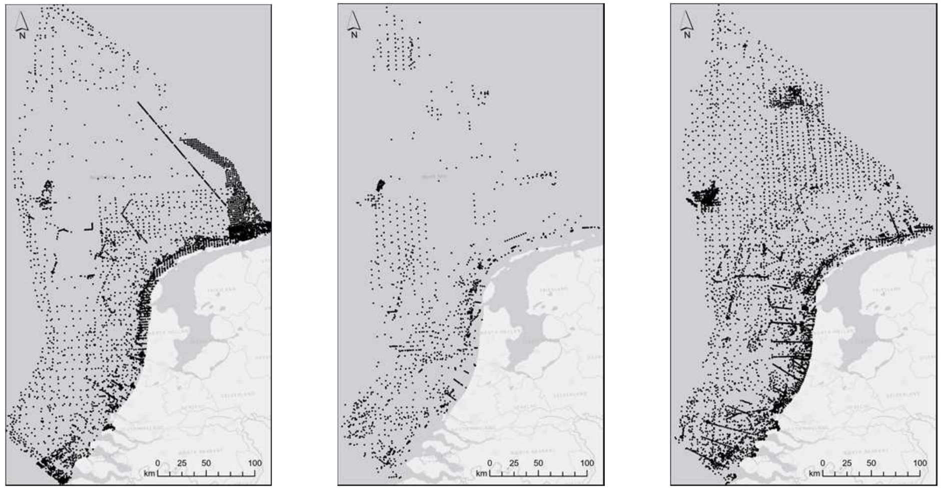
Available data

- Samples collected for half a century (oldest 1934)
- Samples with max. depth 0.5 m below seafloor
  - N=14482
  - If multiple sources at single location: grain size analysis > visual estimation > description
  - Folk sediment class determined for each sample
  - Spatial sampling bias
  - Class imbalance

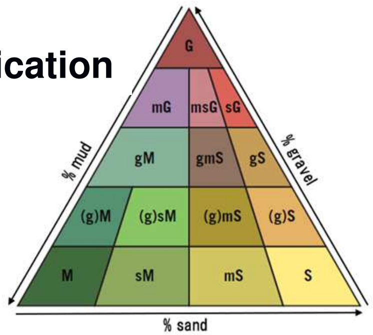
Grain size analysis N=5140

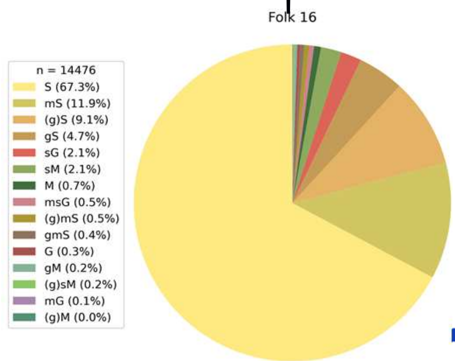
Visual estimation of sediment fraction N=1765

Sediment description N=7577



Folk classification





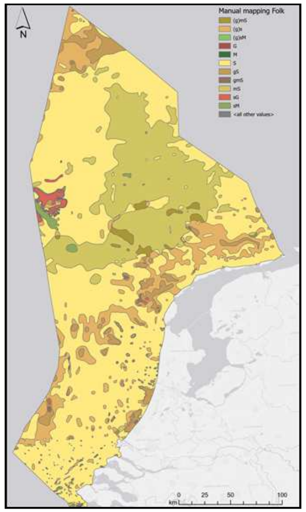
TNO Geologische Dienst Nederland

New map, new methods

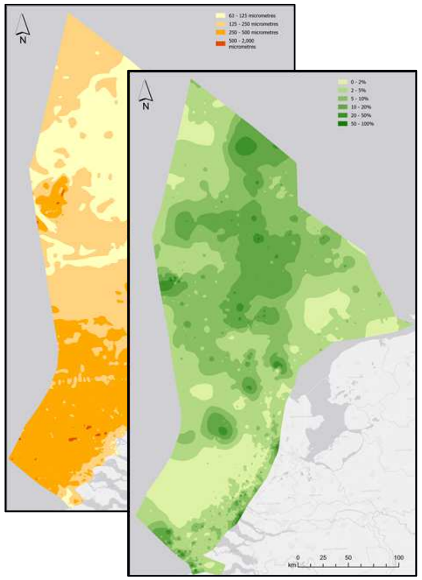
Increasing data density

Increasing mapping speed

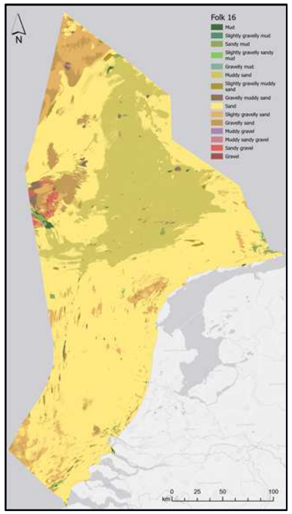
Manual
1980s-90s



Geostatistics
2000s

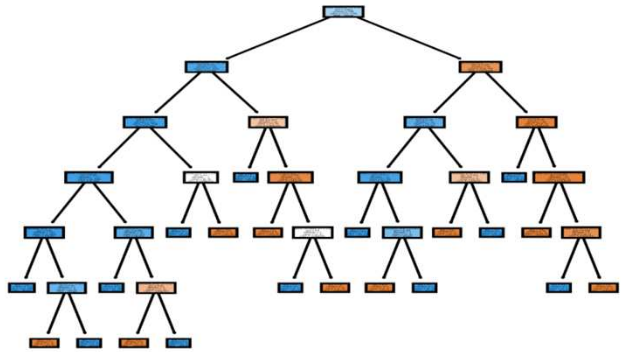


Machine Learning
current

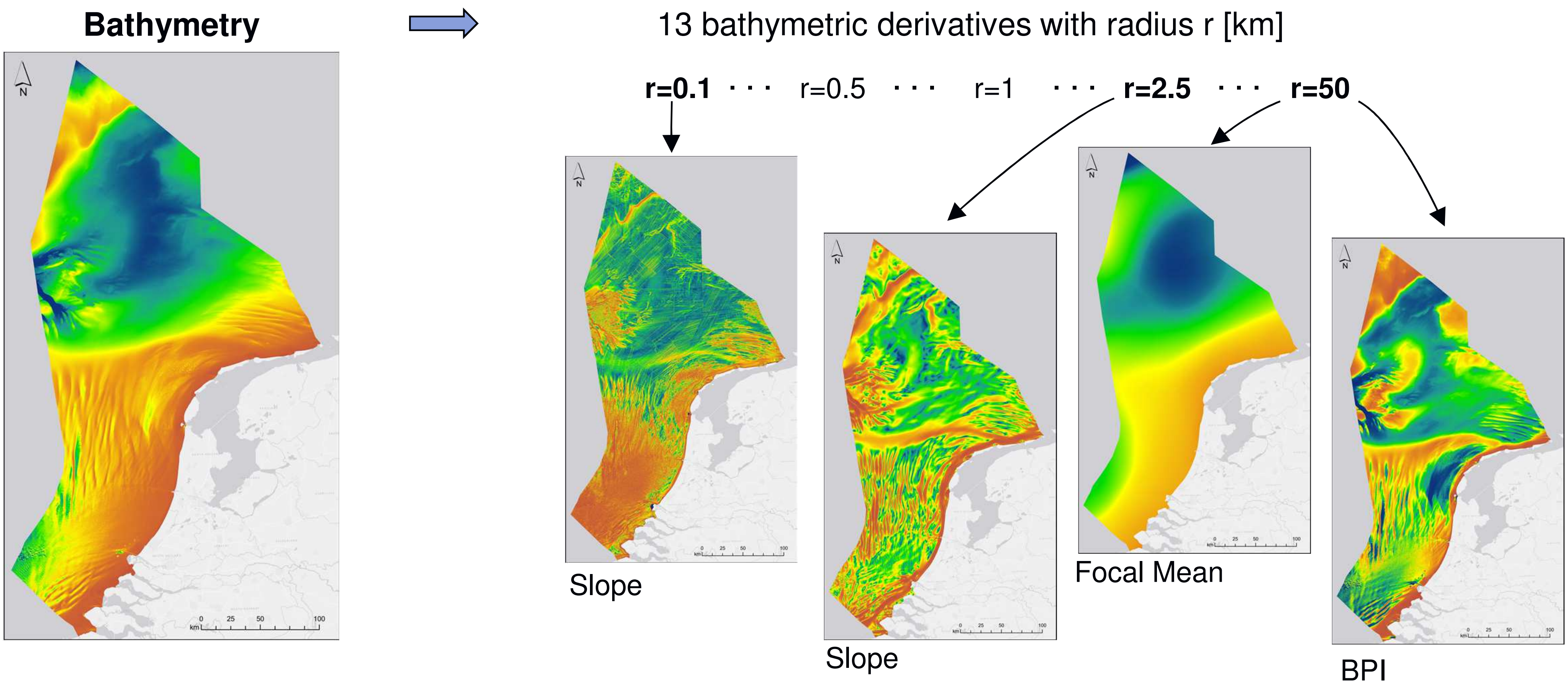


- Expected high inflow of new data in coming years
- Need for method enabling fast future updates

- **ML: Random Forest**
  - Based on decision trees
  - Quick to train & predict
  - Insensitive to hyperparameter tuning
  - Feature engineering → manipulate what it ‘sees’



Feature engineering for Random Forest

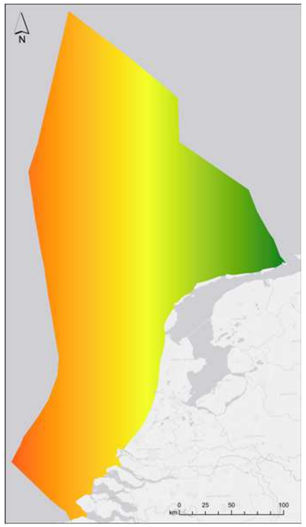


Feature engineering for Random Forest

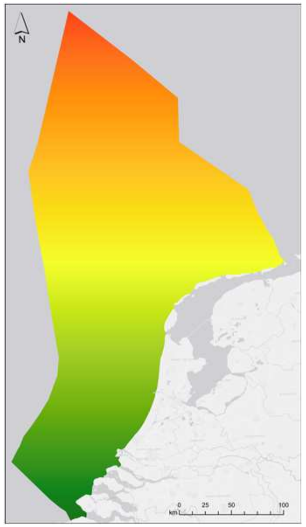
Original coordinates → 12 oblique coordinate* grids for angle α

α=7.5° · · · α=22.5° · · · **α=37.5°** · · · α=52.5° · · · α=67.5° · · · α=82.5°

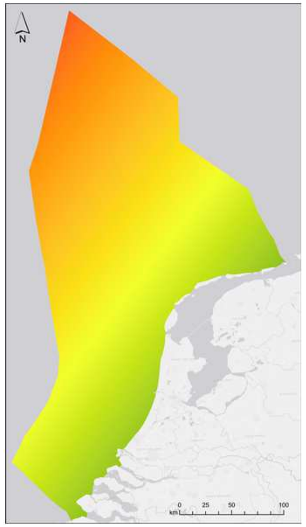
X-crd



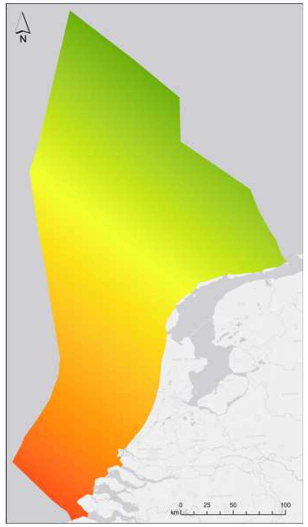
Y-crd



Oblique X-crd



Oblique Y-crd



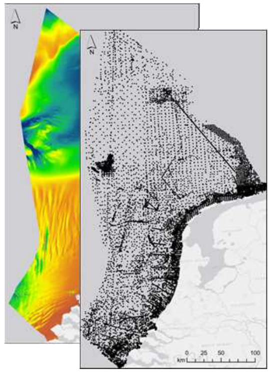
*Møller et al. (2020) Oblique geographic coordinates as covariates for digital soil mapping

Modelling

- Input:
  - Bathymetry
  - Spatial coordinates
- Output: sediment class
- 10-fold cross-validation:
  - Train on 90% of data, predict remaining 10%
  - Repeat 10 times
- Prediction to map
  - Train on 100% of data

At sample locations

Input:
Bathymetry and coordinates



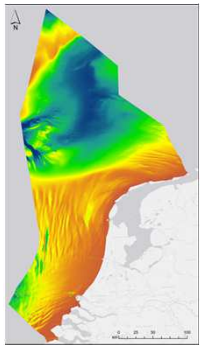
Output:
Folk Sediment Class

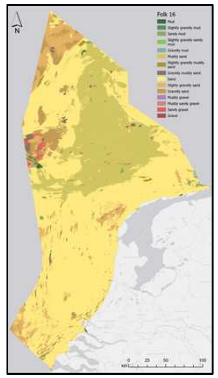
Cross-validation to compare predictions with true values → Cohen's kappa (κ)

ML model

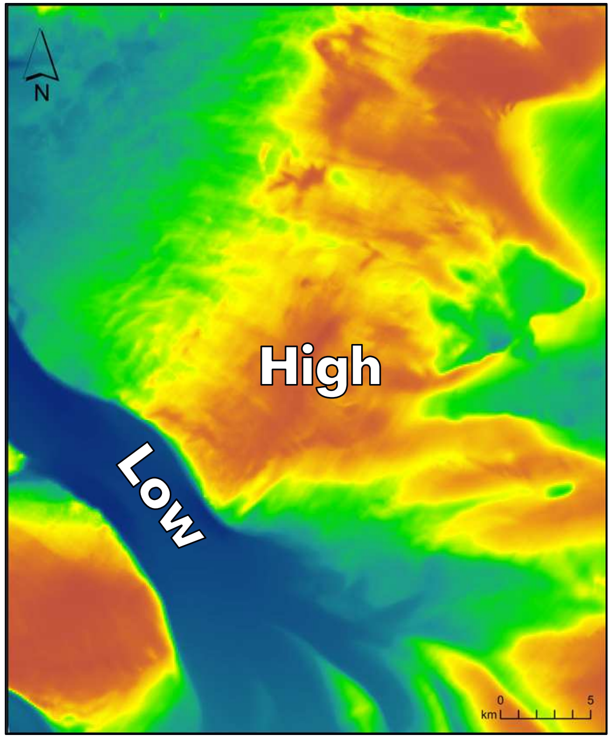
- Predict for every map grid cell
- Visual inspection

At grid cell locations

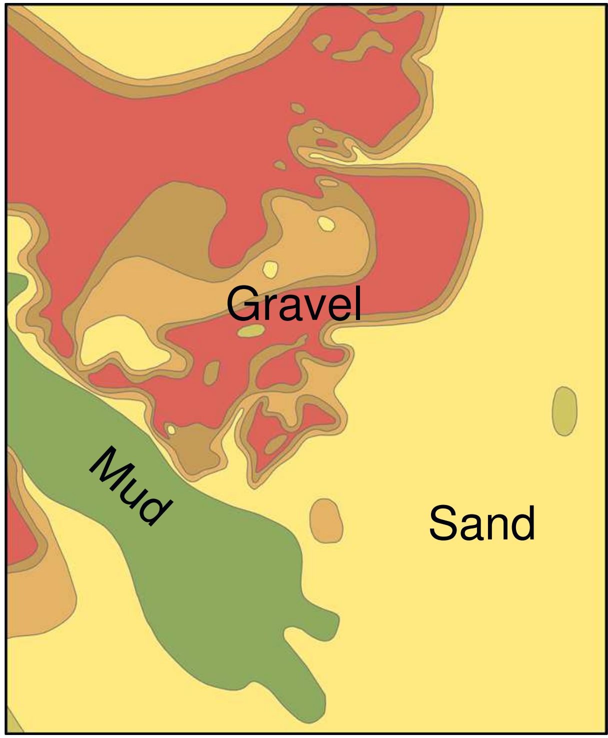




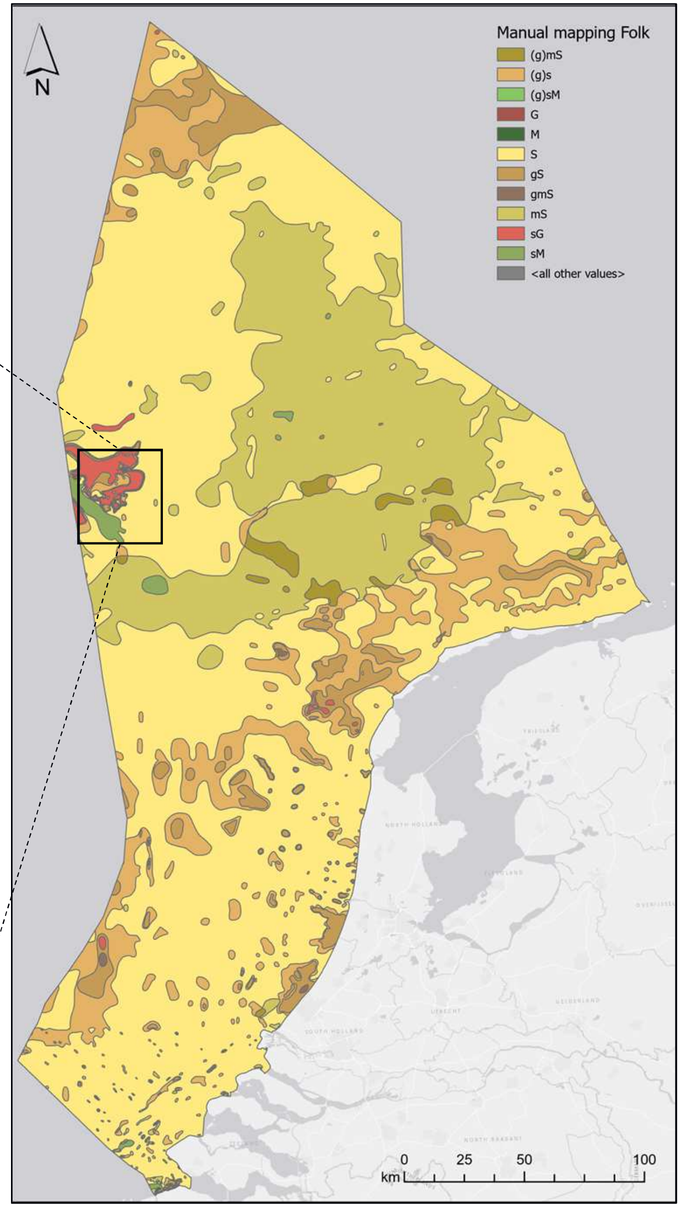
80s-90s Manual mapping



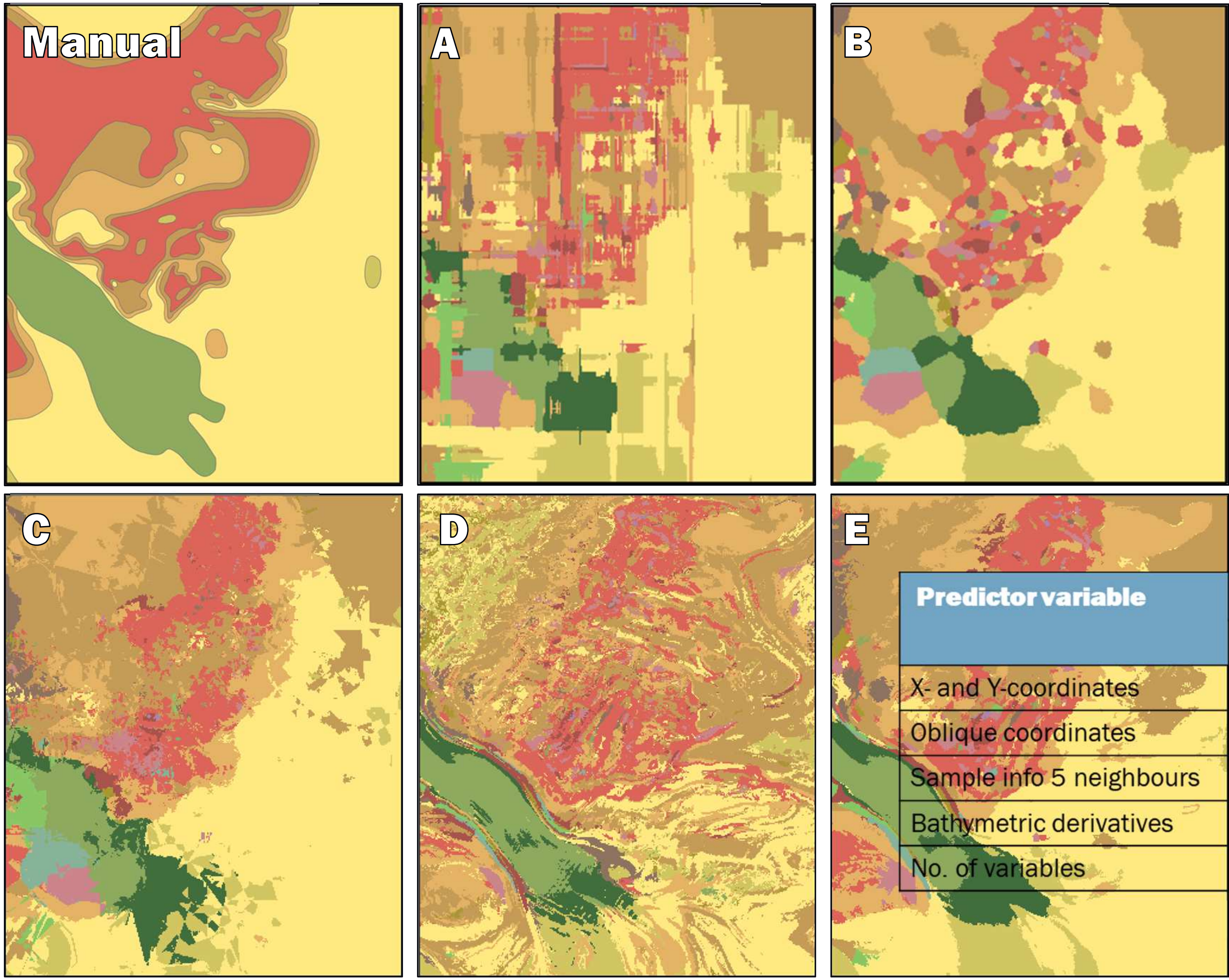
Bathymetry



Folk sediment class



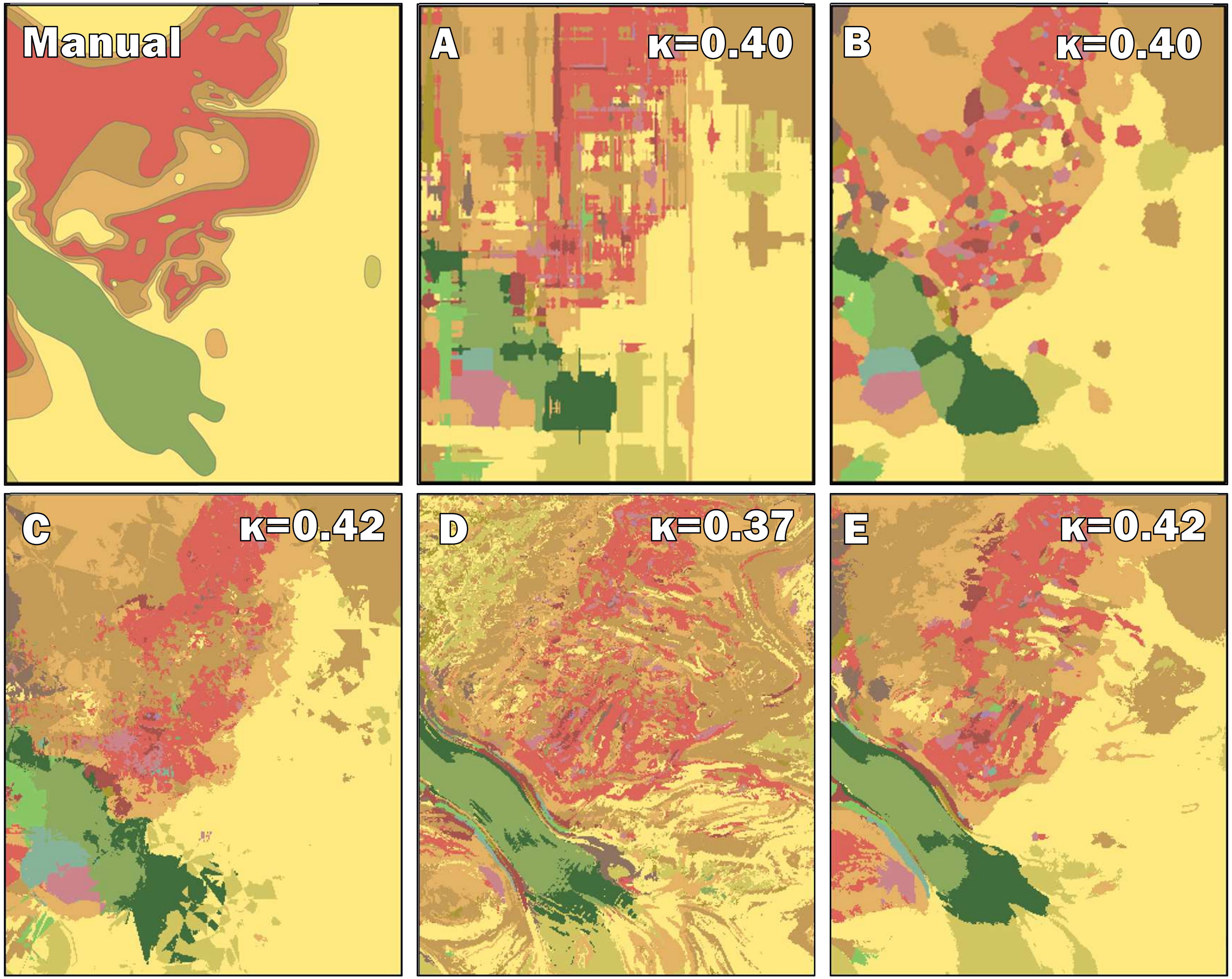
RF Feature selection



Predictor variable	Set				
	A	B	C	D	E
X- and Y-coordinates	x				
Oblique coordinates		x	x		x
Sample info 5 neighbours			x		
Bathymetric derivatives				x	x
No. of variables	2	12	90	13	25

Geologische Dienst Nederland

RF Cross-validation Kappa

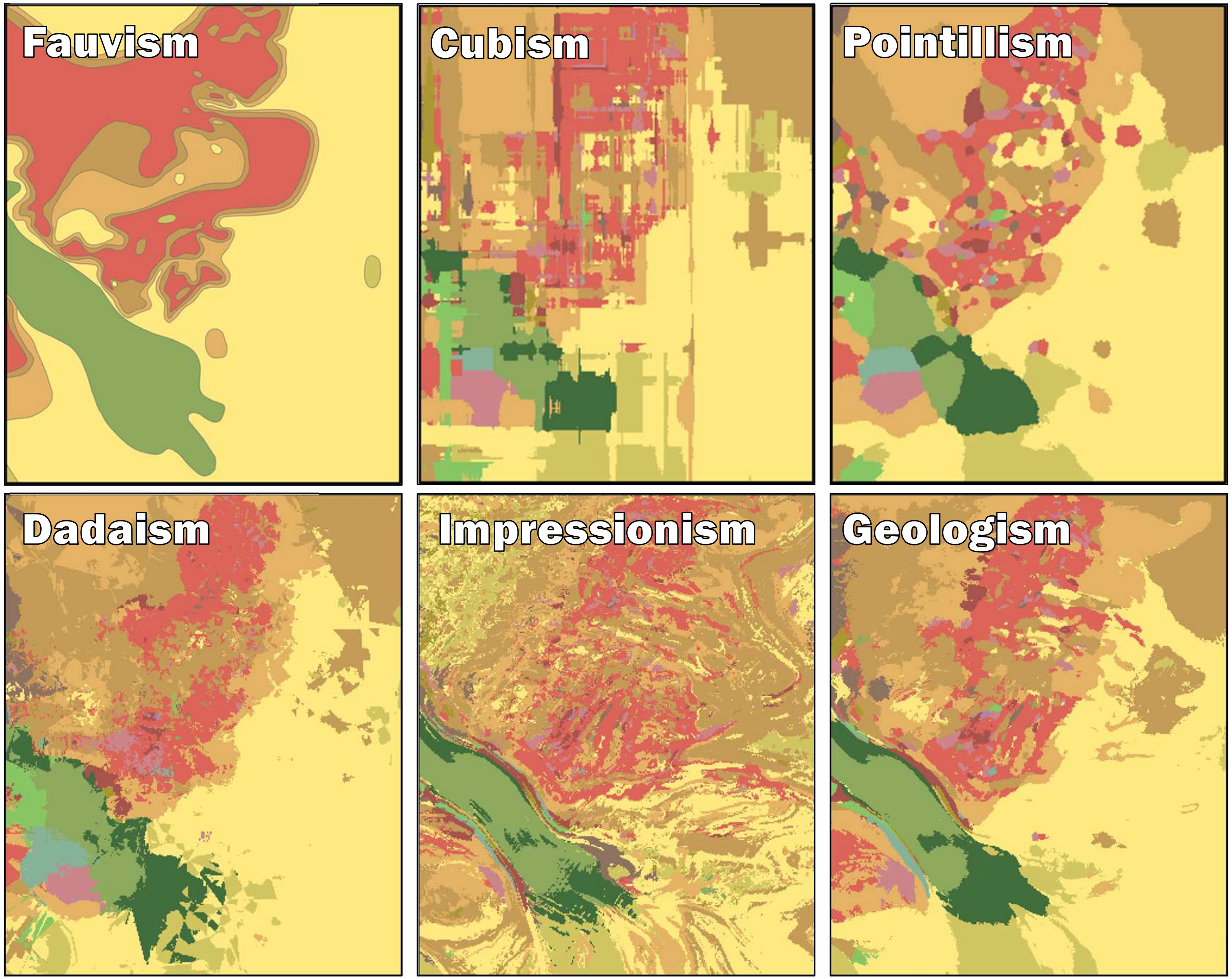


Cohen's kappa

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Geologische Dienst Nederland

RF Seabed-sediment art

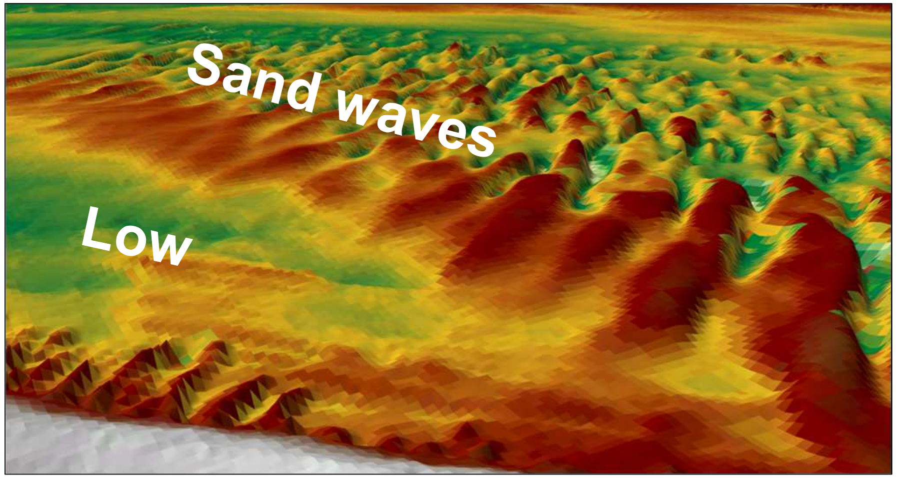


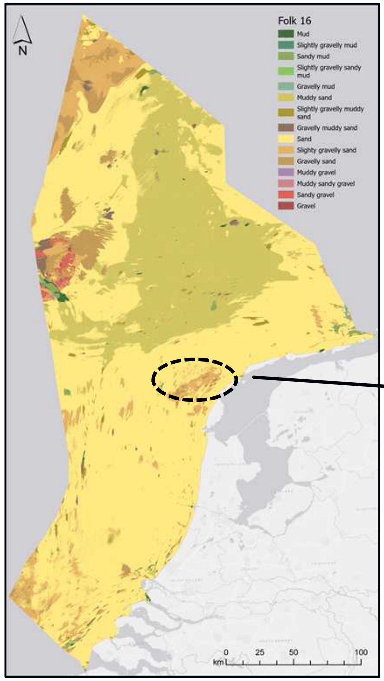
Geologische Dienst
Nederland

Is it geology?

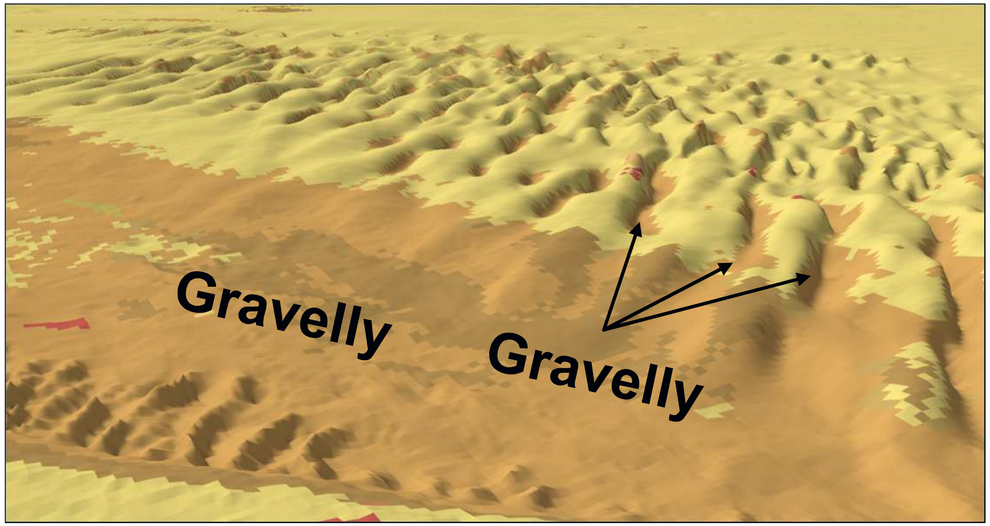
- Geological reality: active sand layer overlying glacial till
- RF Predictions:
  - sand at bathymetric highs
  - coarse sediments at bathymetric lows

Bathymetry



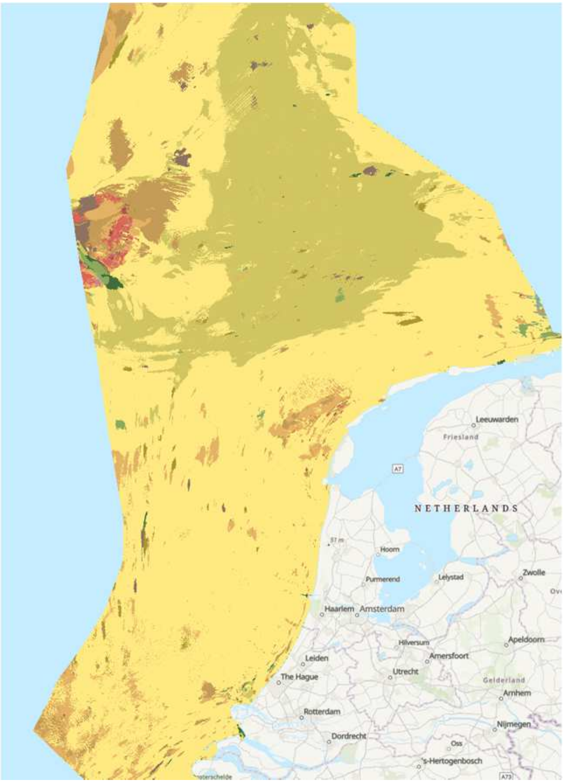


Sediment class

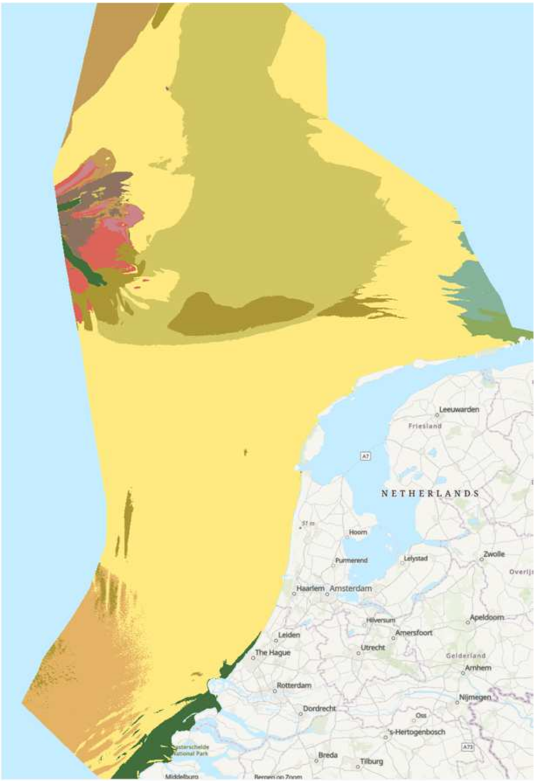


Future: more advanced ML algorithms?

Random Forest

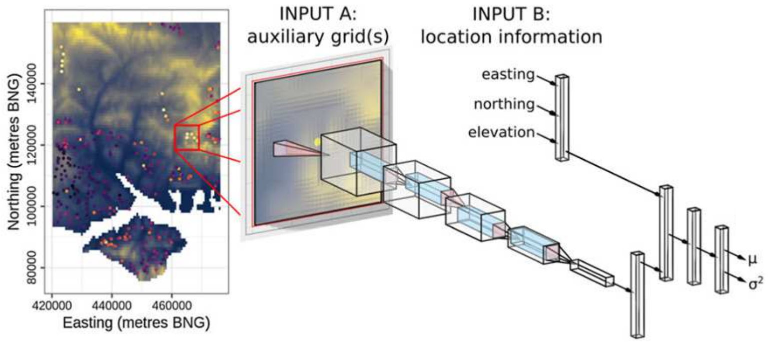


Convolutional Neural Network



Convolutional Neural Network (CNN)

- Neural network to process images
- Takes long(er) to train & predict
- No feature engineering needed
- Hyperparameter tuning and network architecture →
  manipulate how it ‘thinks’



Form: Kirkwood (2022) Bayesian Deep Learning for Spatial Interpolation in the Presence of Auxiliary Information

Geologische Dienst Nederland

Conclusions

What do we have:

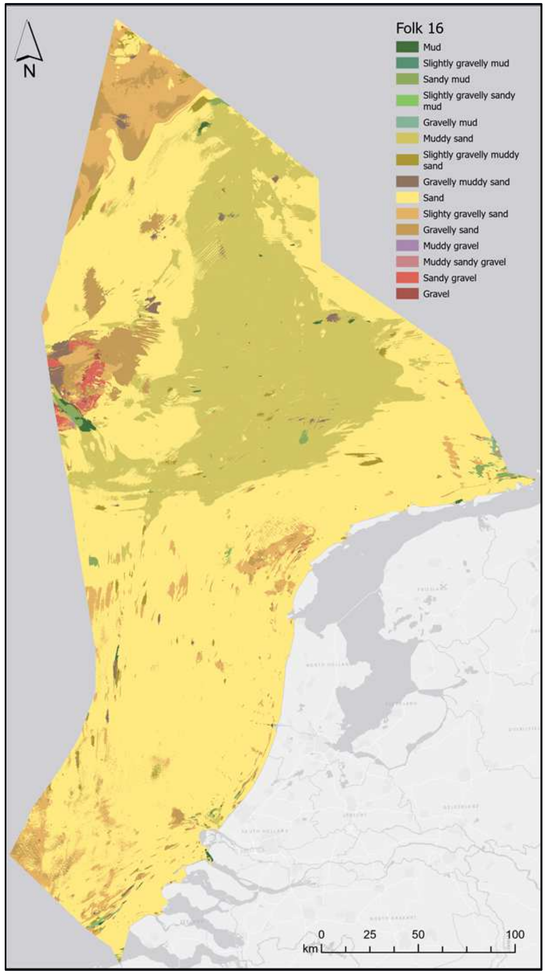
- Seabed-sediment calculated with Random Forest algorithm
  - Based on a lot of data and many bathymetry characteristics
  - Geological complex areas well represented
  - Reliable uncertainty representation
  - Quick to update (minutes)

Possible future improvements:

- Map other sediment characteristics (e.g. grain size, geochemical composition, ...)
- More advanced ML algorithms, e.g. Convolutional Neural Networks
- Combining data driven ML with geological concepts
- ML for all spatial interpolation problems?

THE question:

- Easy to produce many different maps by changes in input features or ML algorithm. But what is the best map? How to quantify geological plausibility?



TNO Geologische Dienst Nederland