IMPERIAL

In collaboration with IIT Kanpur

Towards Better Sand Mining Monitoring:

Generalisation or memorisation in machine learning (for remote bathymetry derivation)

Raul Adriaensen
PhD in Coastal Modelling
Raul.adriaensen17@imperial.ac.uk

Supervisors:
Professor Matthew D. Piggott
Dr. Philippa J. Mason
Dr. Yves Plancherel

In collaboration with:
Shobhit Sing
Professor Rajiv Sinha

Structure

1. Background and Case Study
2. Achieving Optimal Performance
3. Minimum Field Data Requirements
4. Conclusions

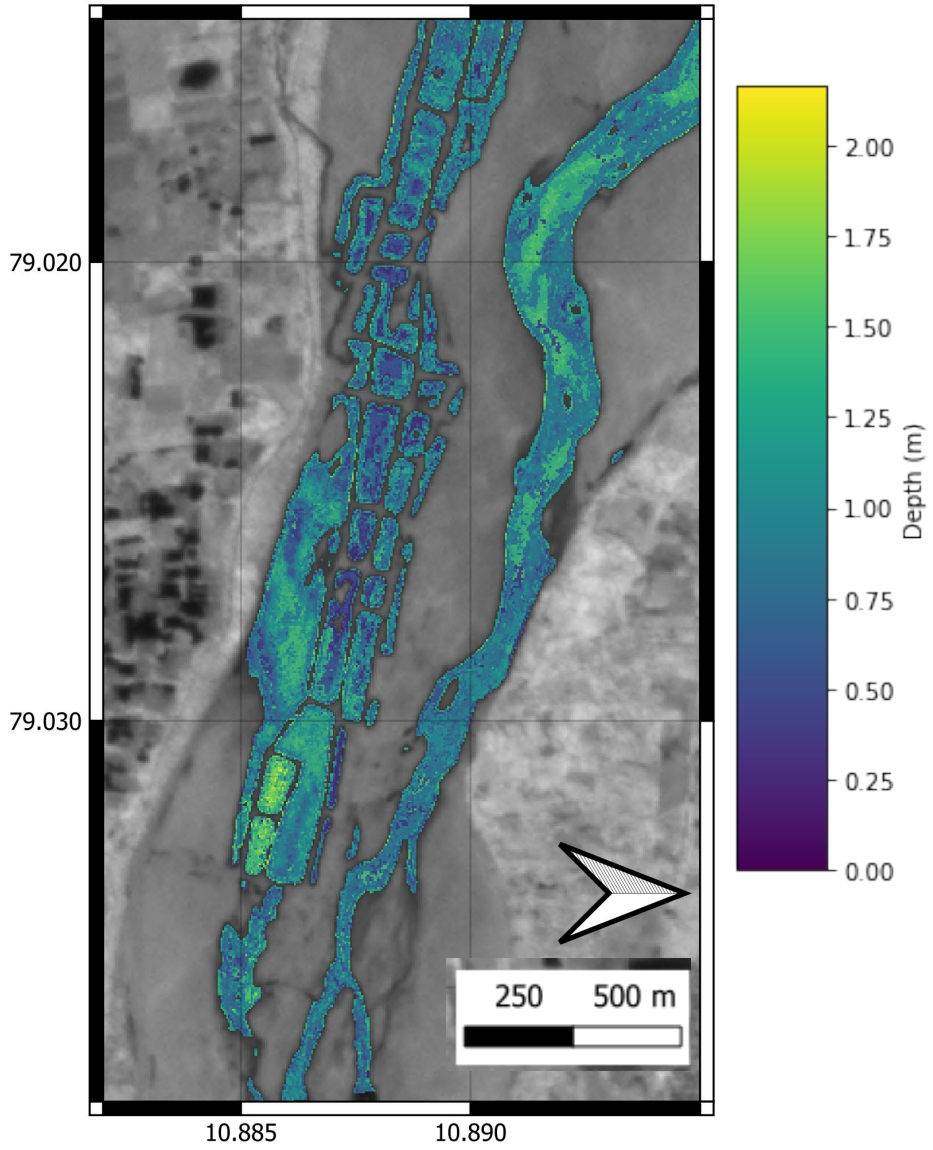


Fig 1. Satellite Derived Bathymetry product superimposed on Drone/Sat data

1. Background and case study

1. Background and Problem

PhD context

Dominant error source in coastal hydrodynamic modelling is bathymetry

1. Updating Maldives/Chagos bathymetry products
2. Unstructured meshing as a function of bathymetry
3. Shoreline structure classification
4. Updating tidal boundary conditions by assimilating bathymetry, tide gauges and altimetry data (SWOT)
5. Flow tracking for validation

Satellite Derived Bathymetry (SDB)

1. Synthetic toolbox to generate optical imagery for the study and standardisation of SDB algorithms.

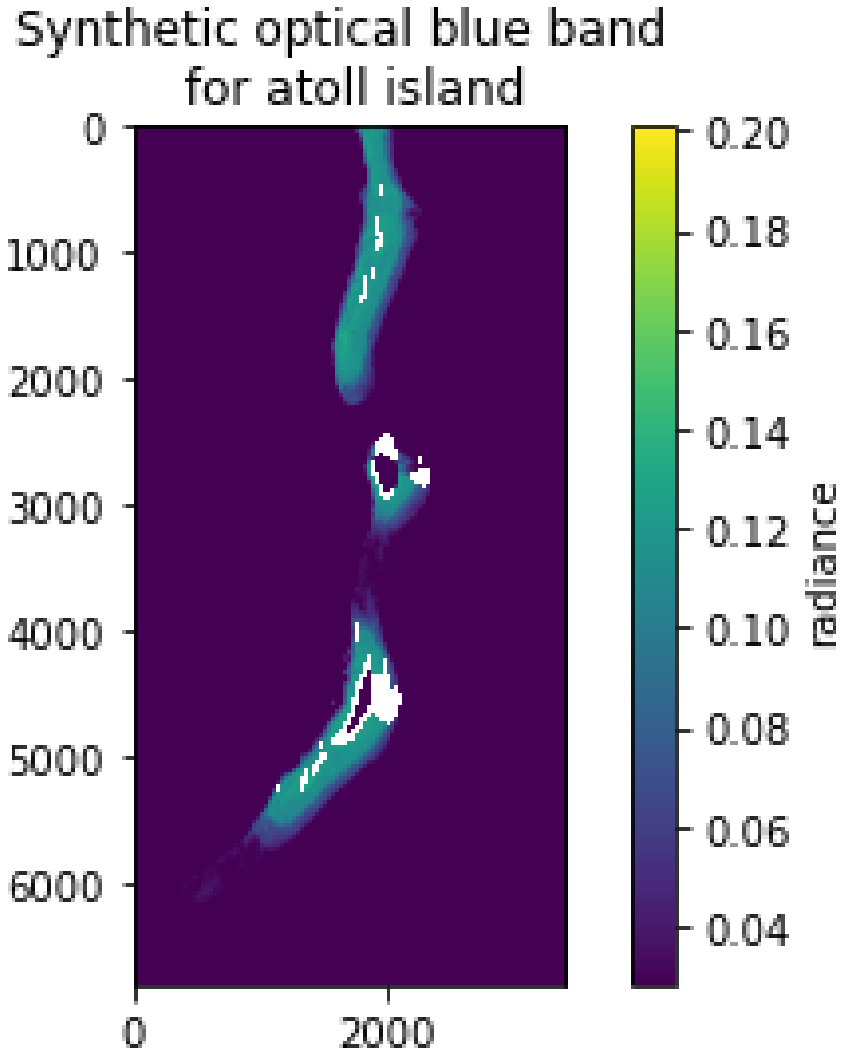


Fig 2. Synthetic Optical image generated with our toolbox, replicating bathymetry of Chagos atoll edge with assumed white sand everywhere. Cloud and land masked

1. Background and Problem

Literature: What is SDB

- Empirical V Physics
- Performance to ~20 meters
- Radiance values is primarily a function of:
  - Water column properties (Absorption & Backscatter)
  - Bottom Albedo
  - Depth

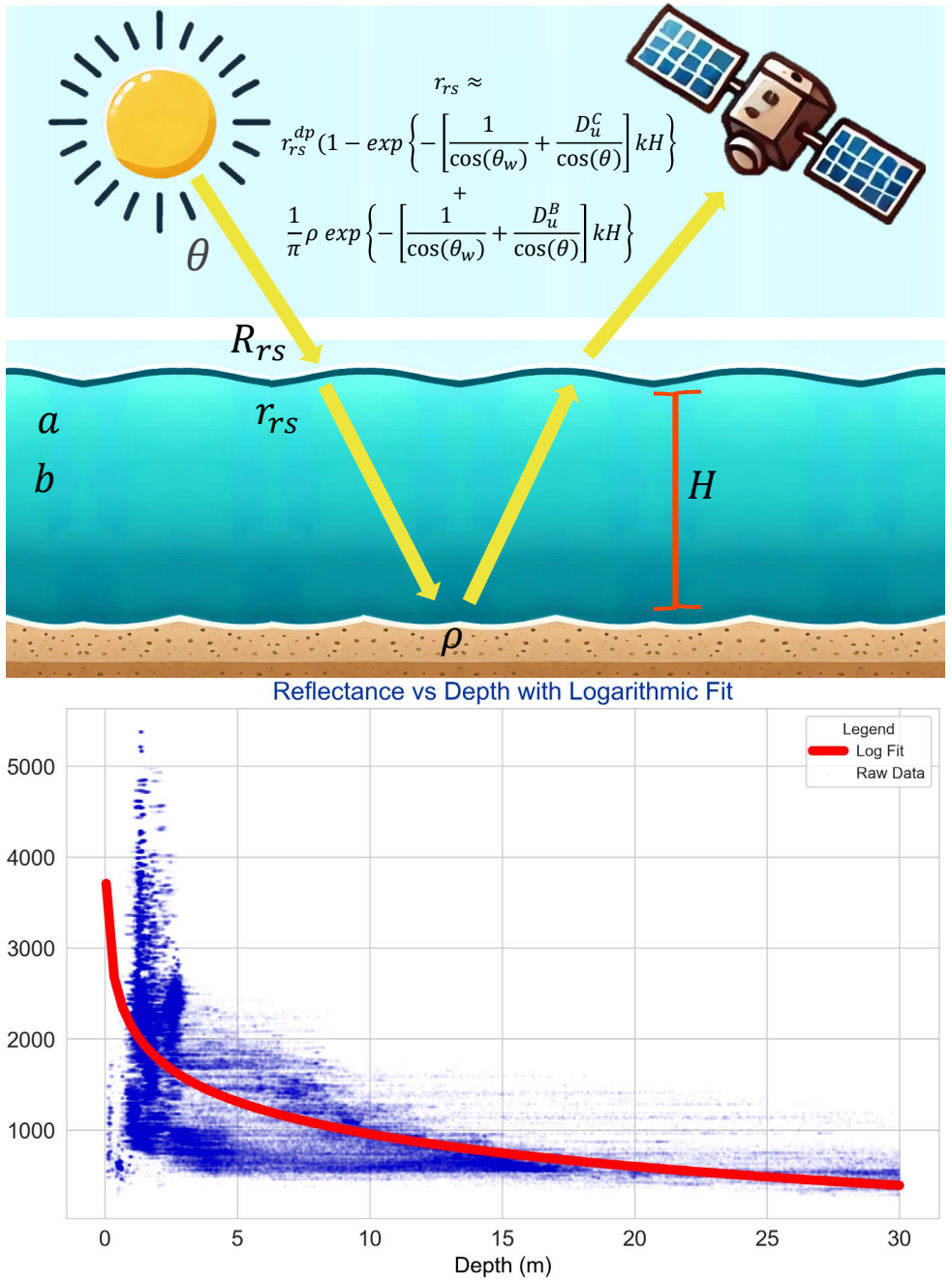


Fig 3. Empirical versus Physics based methods approach. Top diagram representing radiative transfer equations, bottom showing relationship between reflectance and depth with real data

1. Background and Problem

Case Study: Illegal Sand Mining

1. Majority of sand mining concession are issued legally, but operators frequently exceed permissible volumes by an order of magnitude causing.
    - > economic losses
    - > environmental damage
    - > loss of resources (useful sand is finite!)
2. Initial monitoring strategies included drone DEM creation to identify mining pit volumes
3. Method no longer feasible due to miners now infilling pits with water to obscure illicit activities, perfect for SDB!
4. Important to sue people fairly!

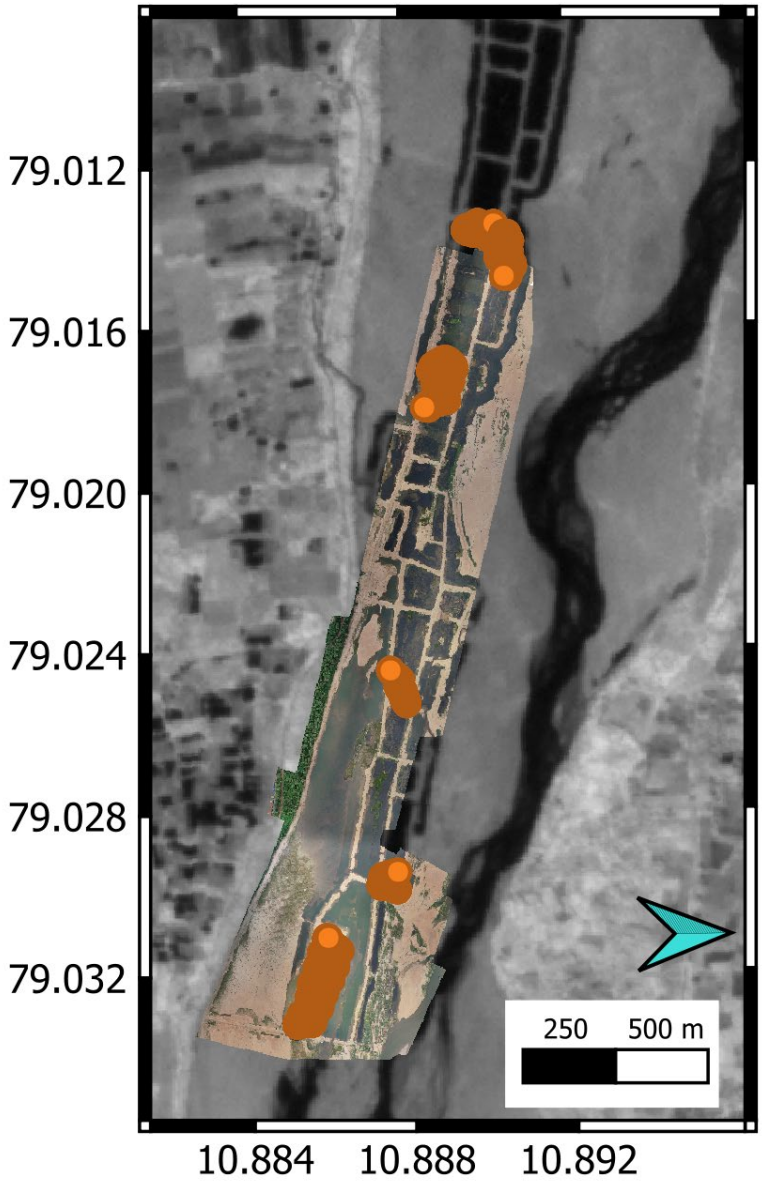


Fig 4. Drone mosaic on top of Planet infrared image of Maruvur locality visualising the sand mining activity occurring and bathymetry data

Case Study Aims

1. Achieving Optimal Performance

2. Understanding minimum required data collection efforts

2. Achieving Optimal Performance

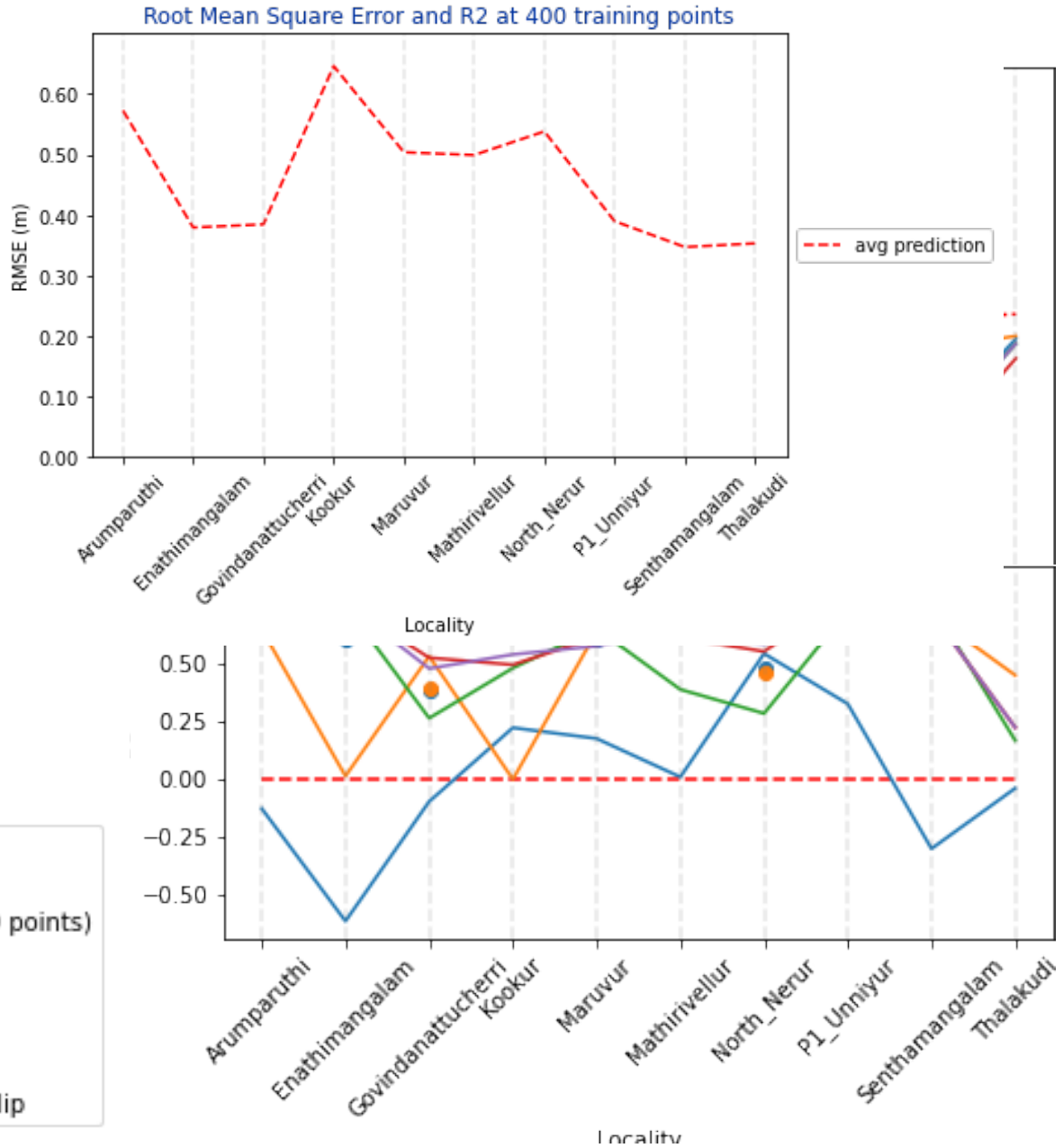
What is optimal?

Achieving Optimal Peformance

Preprocessing steps

Test effect of each step:

1. Raw bathymetry
2. **Cut statistical outliers**
3. Automated and manual land removal
4. Depth measurements averaging per pixel
5. Features in such as MNDWI
6. Cut shallowest 0.2 m (interference rover)
7. Varying input features

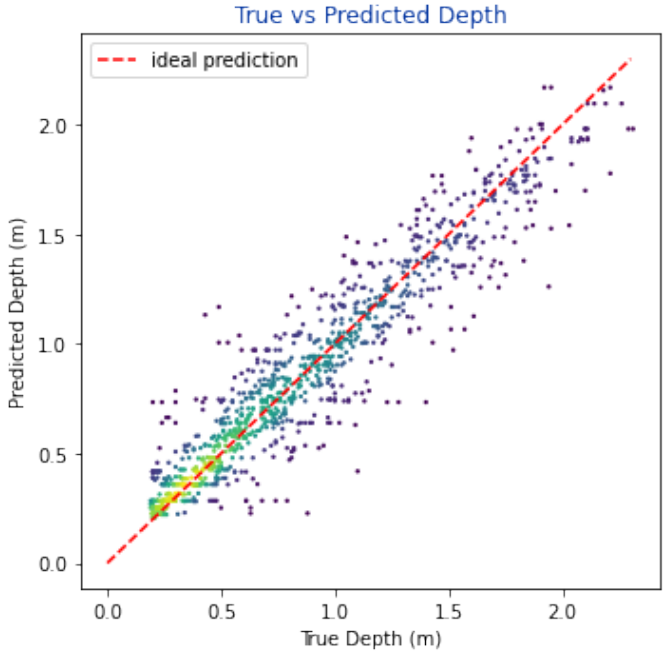


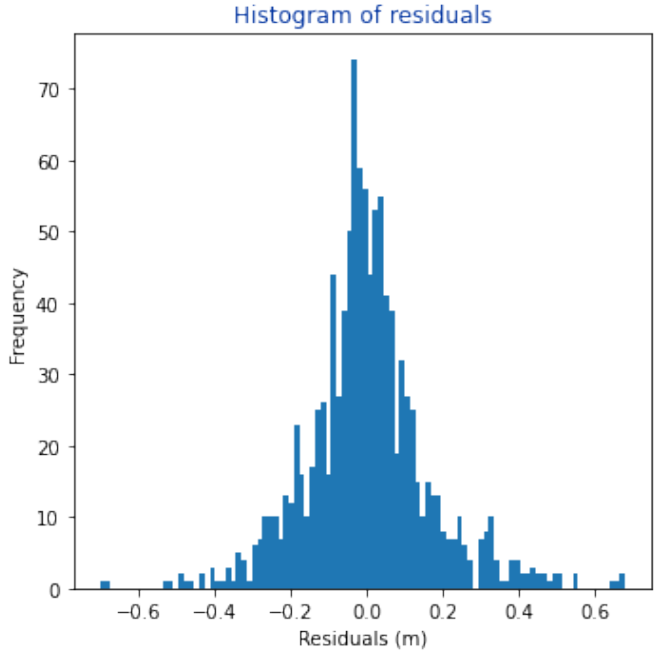
Achieving Optimal Performance

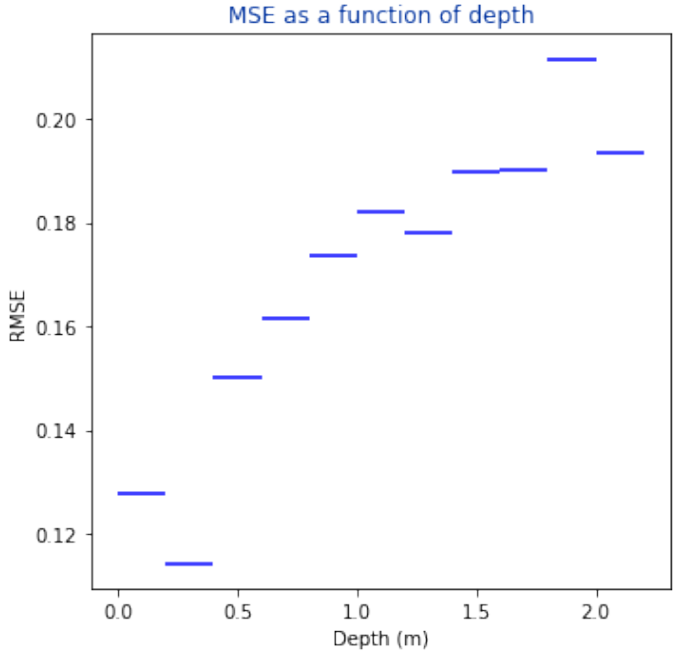
Hyperparameter tuning

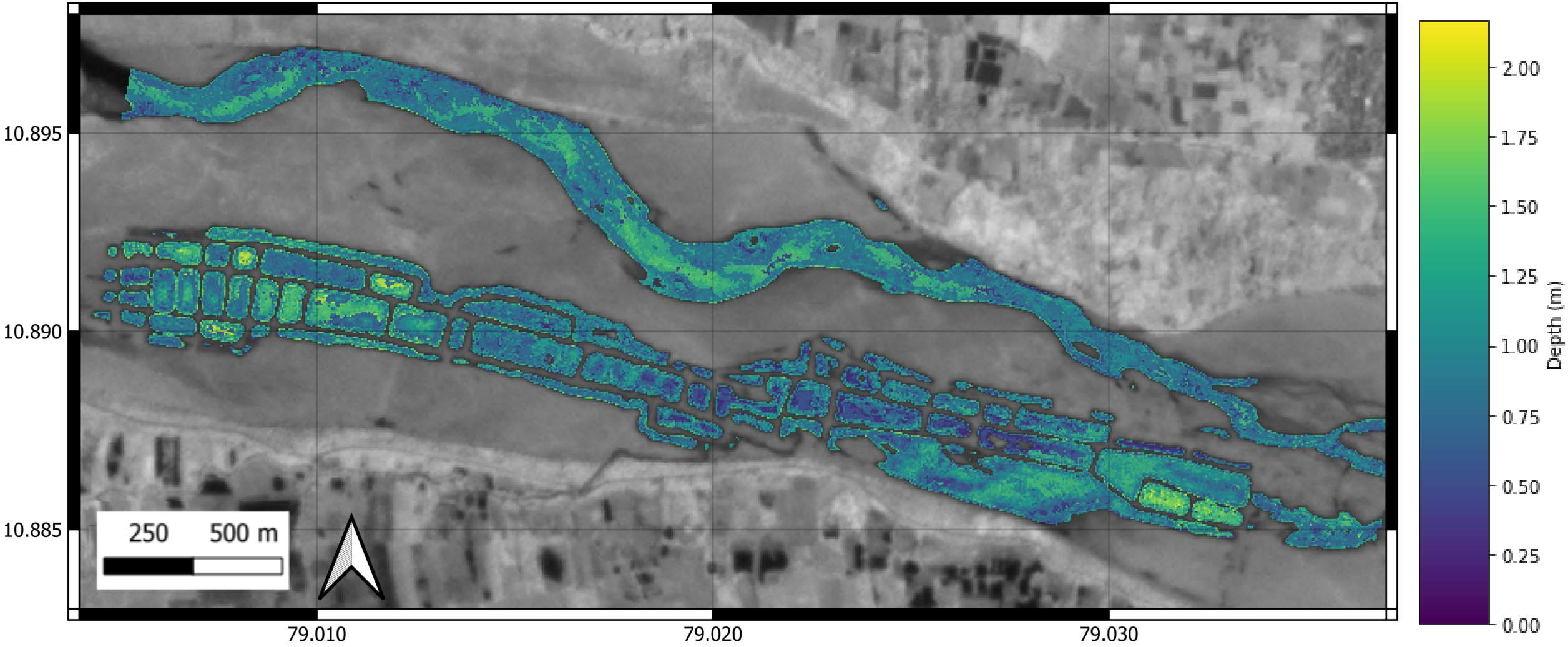
- K-NNs
- 50/50 train/test **random** data split
- RMSE of 0.16 m and R2 of 0.9

Interpolating the data we get an RMSE of 0.1m





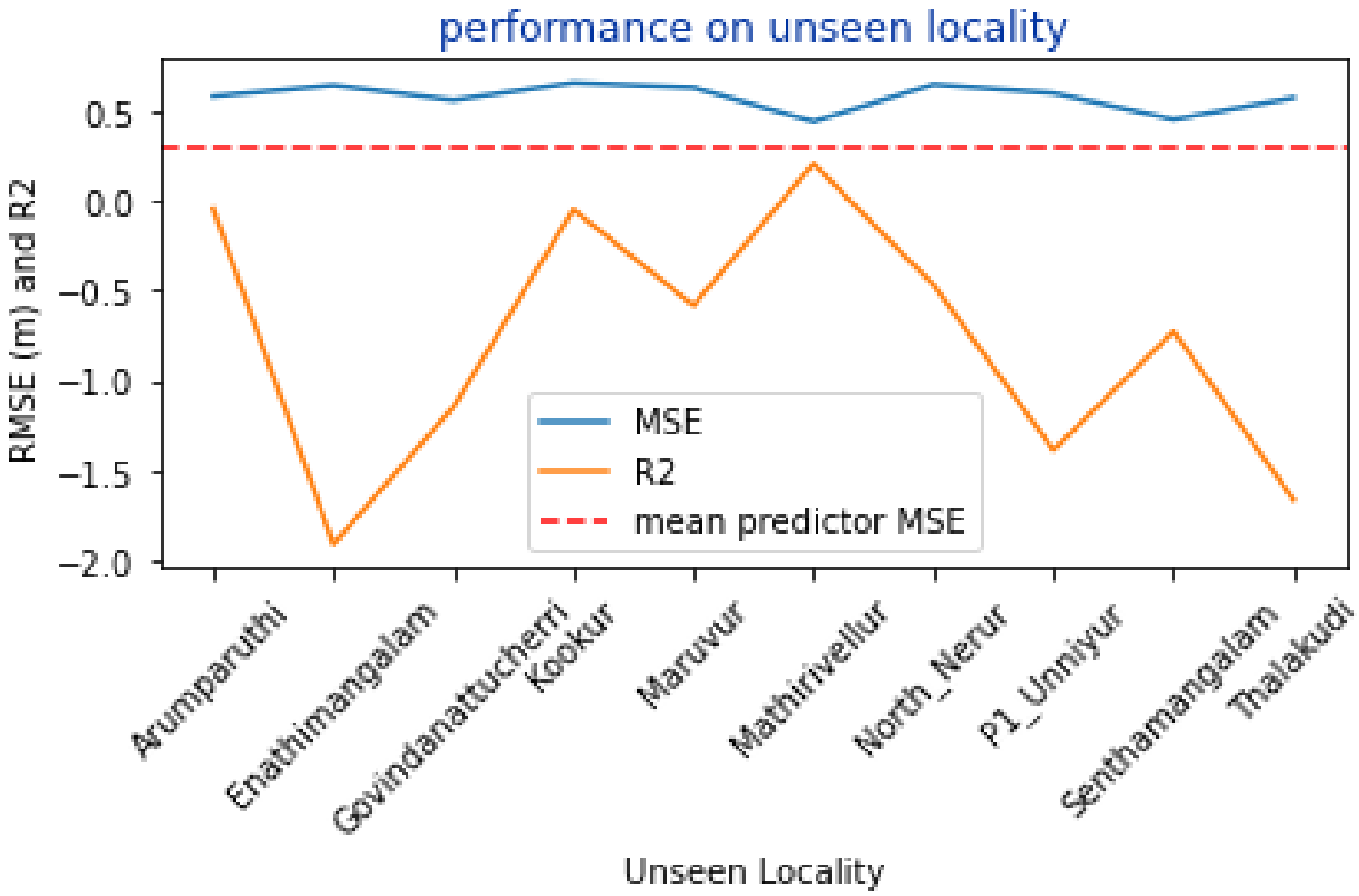


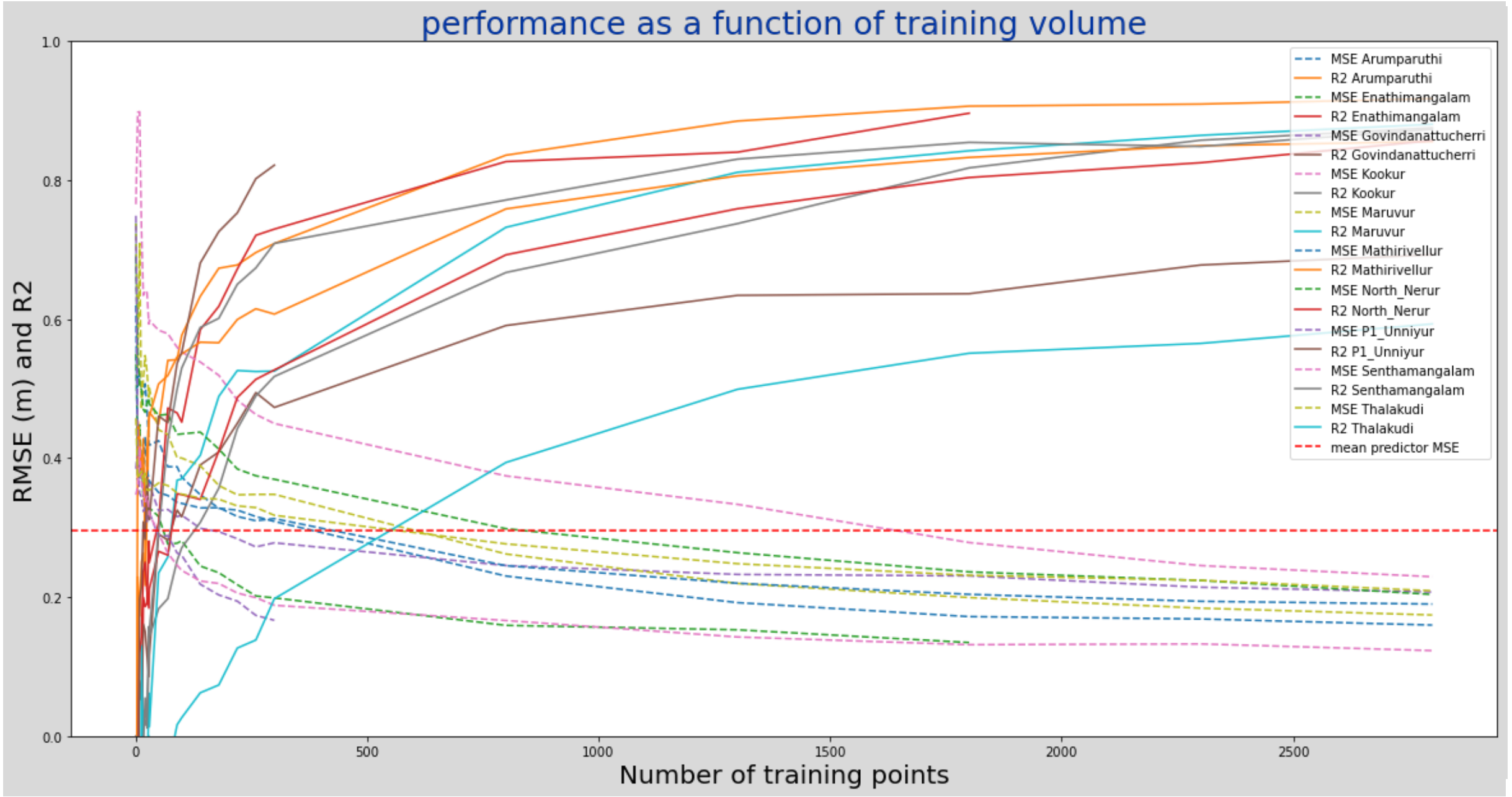

10.895
10.890
10.885
79.010
79.020
79.030
250
500 m
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Depth (m)

3. Minimum data requirements

Minimum data requirements

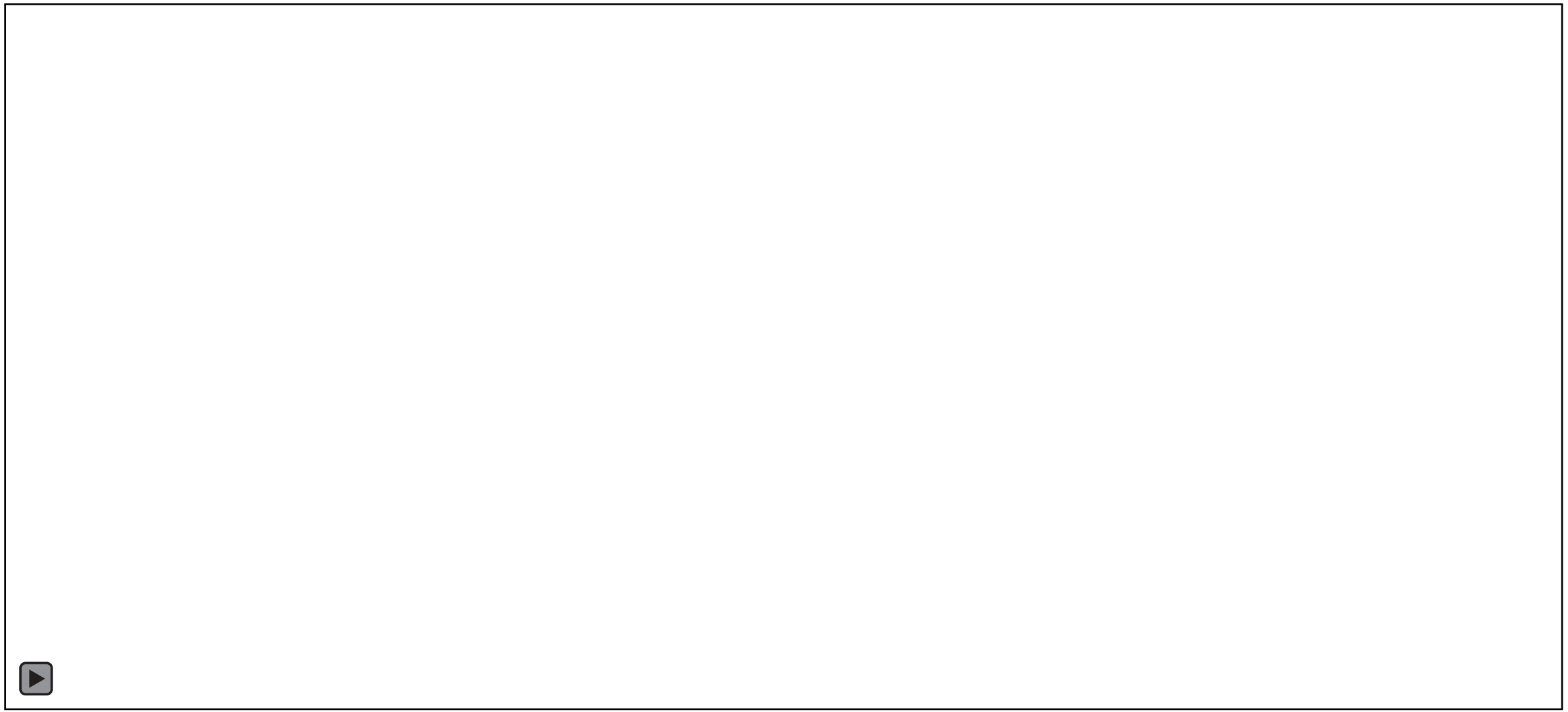
Do we need in situ data?




performance as a function of training volume
RMSE (m) and R2
Number of training points
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
MSE Arumparuthi
R2 Arumparuthi
MSE Enathimangalam
R2 Enathimangalam
MSE Govindanattucherri
R2 Govindanattucherri
MSE Kookur
R2 Kookur
MSE Maruvur
R2 Maruvur
MSE Mathirivellur
R2 Mathirivellur
MSE North_Nerur
R2 North_Nerur
MSE P1_Unniyur
R2 P1_Unniyur
MSE Senthamangalam
R2 Senthamangalam
MSE Thalakudi
R2 Thalakudi
mean predictor MSE

Minimum data requirements

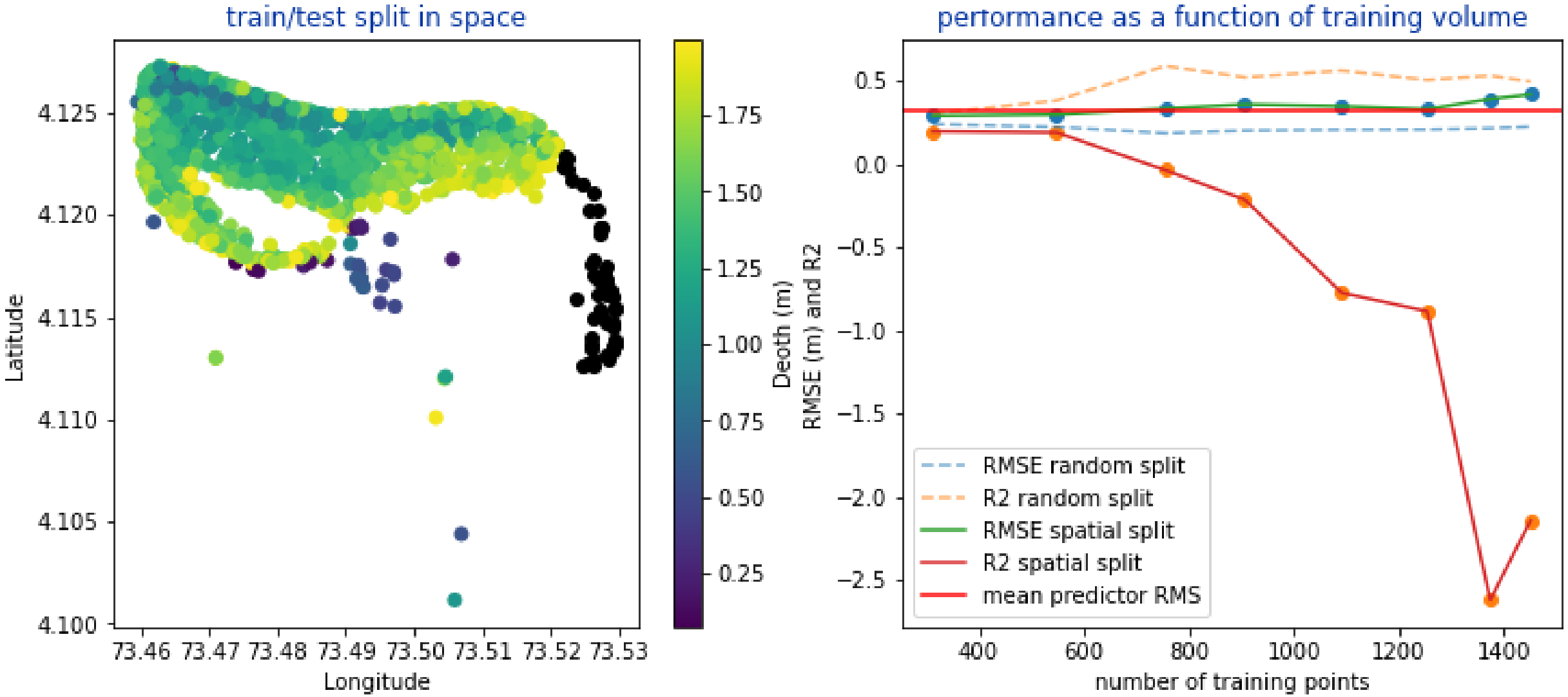
Generalise in one locality



Minimum data requirements

Data comparison

0 – 2 m

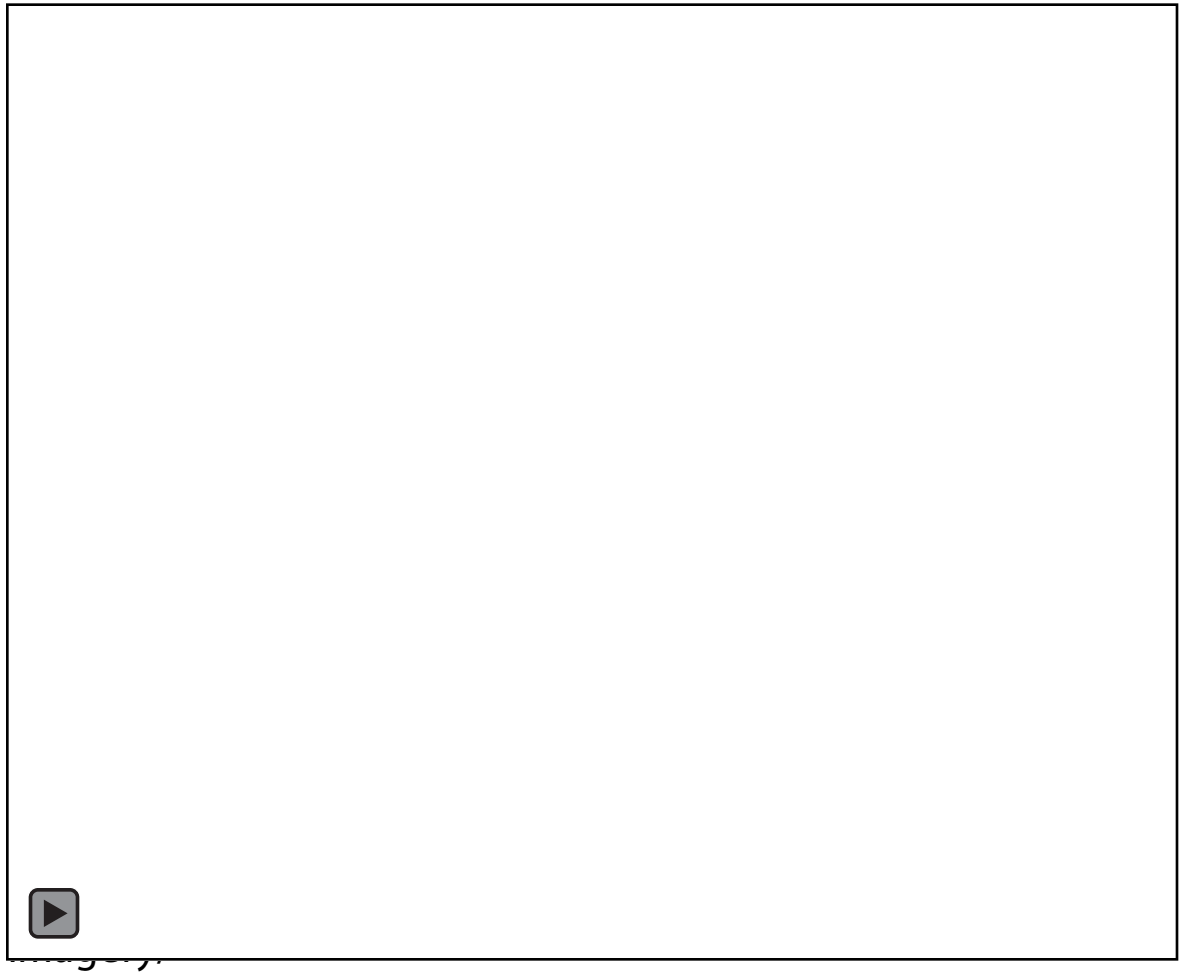


Conclusions and Future Work

1. Sampling strategy is key to demonstrate true performance
2. Need a baseline performance measure to give errors measures meaning. For us either interpolation or predicting the mean
3. The extremely shallow bathymetry is riddled with complexity in both coastal and river

Next work for policy recommendation:

- Additional data collection
- Automated identification of sand mining activity
  - Hopefully with RAIC (Rapid Automated Image Detection) tool!



IMPERIAL

Thank you

Towards Better Sand Mining Monitoring: Tackling Memorization in Machine Learning for Remote Bathymetry Derivation

12/12/2024