Towards a paradigm shift on mapping muddy waters in water reservoirs with Sentinel-2 using Machine Learning

MKLab
M4D
MultiMoDal Data Fusion and Analytics Group
https://m4d.iti.gr

Konstantinos Vlachos, Christos Psychalas, Anastasia Moumtzidou, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris

{kostasvlachosgrs, psychalas, moumtzid, heliasg, stefanos, ikom}@iti.gr

Information Technologies Institute, Centre for Research and Technology Hellas
6th km Harilaou - Thermi, 57001, Thermi, Thessaloniki, Greece

CENTRE FOR RESEARCH & TECHNOLOGY - HELLAS
Information Technologies Institute

WQeMS

1. Intro

Heavy rainfall comprises a rising **extreme event** due to climate change, which induces increased run-off, high-flow, **flooding** and **erosion** leading to increased sediment particles in the water. The erosion is especially high when soil is exposed in regions such as construction/mining sites, burned areas, or areas with poor agricultural/forestry practices, which in turn results in **large volumes of sediment** entering the water abruptly, constituting it as "**muddy**". Some **negative implications** are on i) a **lake ecosystem** (e.g., altering the physicochemical regime) ii) **human infrastructure** (e.g., dams, water utilities hardware) and iii) the **local economy** (e.g., when used for recreational purposes), hence the importance to monitor with **high spatial resolution** and **frequency**.

2. Methodology

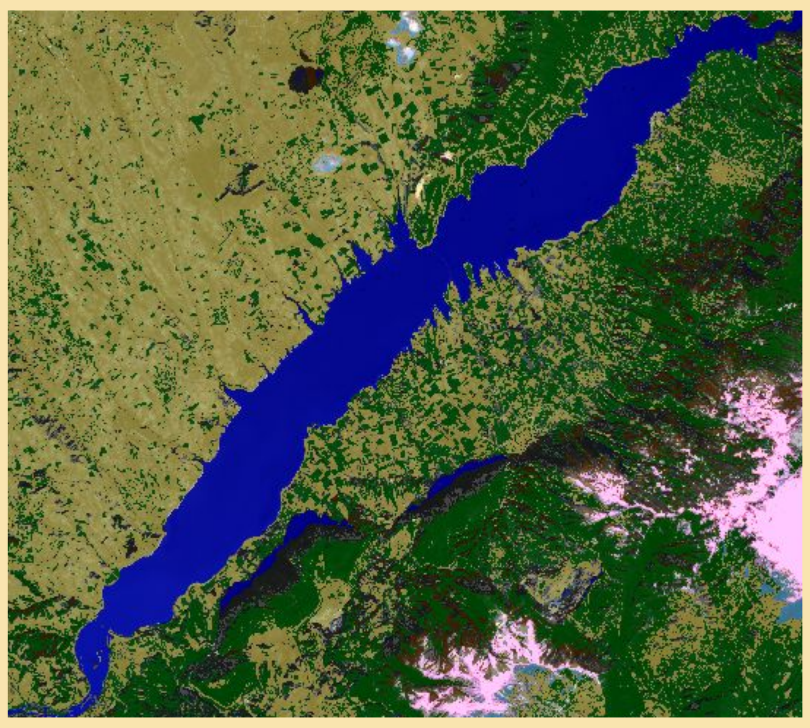
Data:
- Sentinel-2 Level-1 (10m, ~5 days)

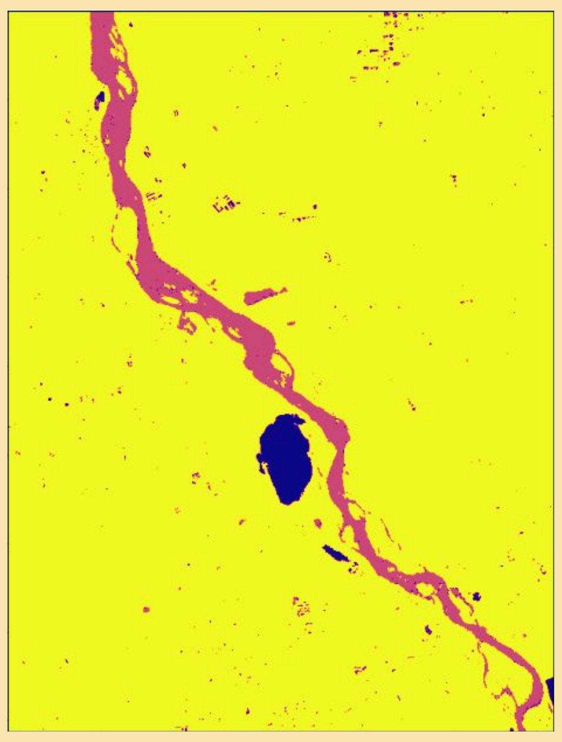
Annotation:
- Scene Classification Layer (**SCL**)
- Combination between **NDWI** and **NDTI**
- Expert knowledge based on **RGB** / **FCCs**

Preprocessing:
- Select various scenes (Greece, Finland) with inland / coastal muddy water presence
- Apply Atmospheric Correction (Sen2Cor)
- Resample to 10m (bilinear interpolation)
- Subset / Rescale

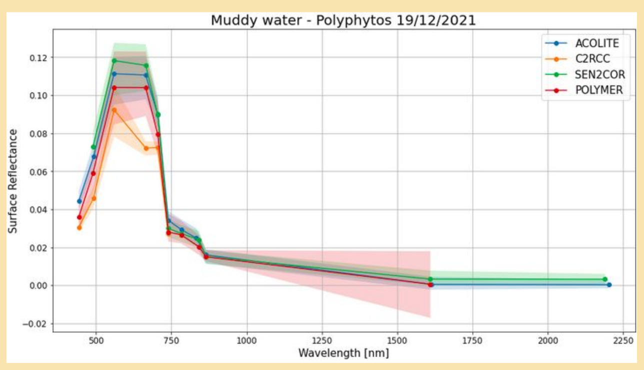
Model / Training:
- Input: Bands → B{2-8}, B8A, B{11-12}
- Random Forest model
- Samples:
  - Total: 166738
  - Training: 133390
  - Test: 33348
- Hyper-parameter tuning based on:
  - Grid Search:
    - Tree depth [4, 5, 6, 8,10, 20, 30, 60, **100**, 120, 150]]
    - Number of trees [5, 6, 7, 10, 20, 50, **100**, 150, 200]
    - Class balance with weights [**sample balance**, none]
  - 5-Fold Cross Validation with 5 repeats
  - F1-score
  - Total number of fits:7425

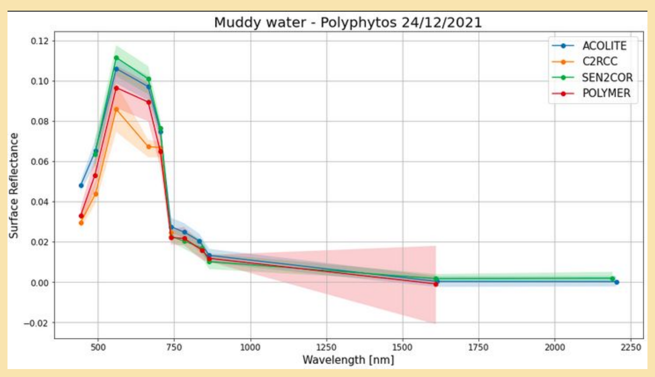
3. Analysis & Evaluation

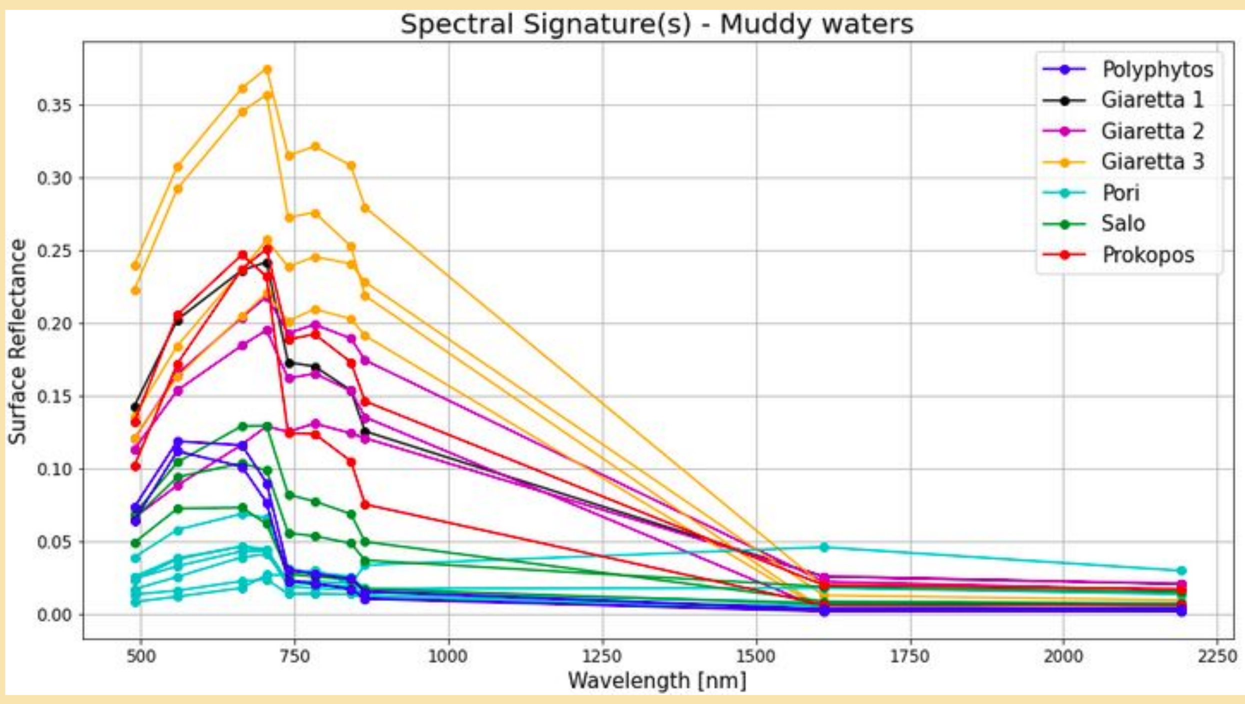

SCL


NDWI / NDTI


RGB



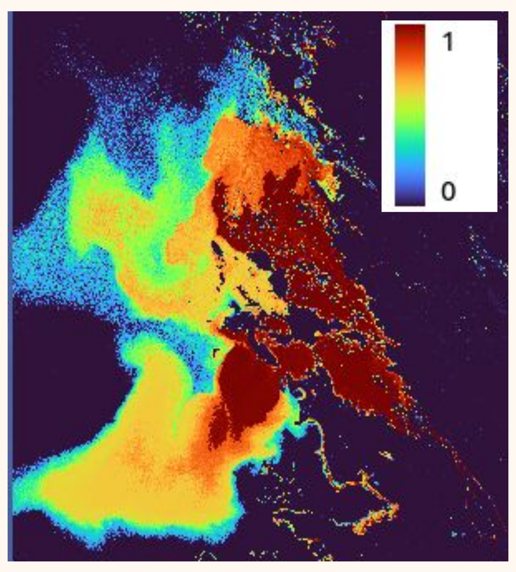


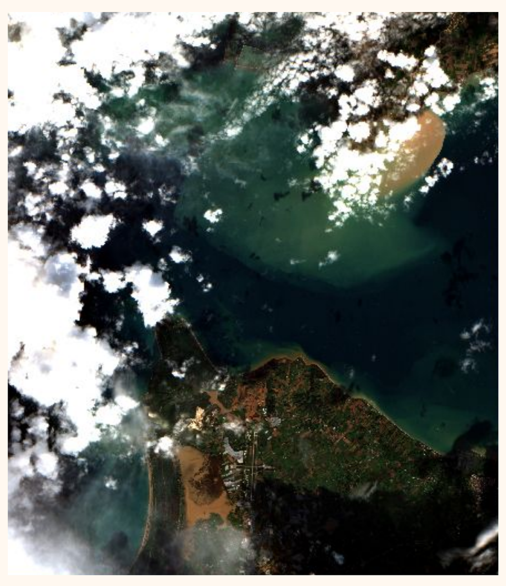


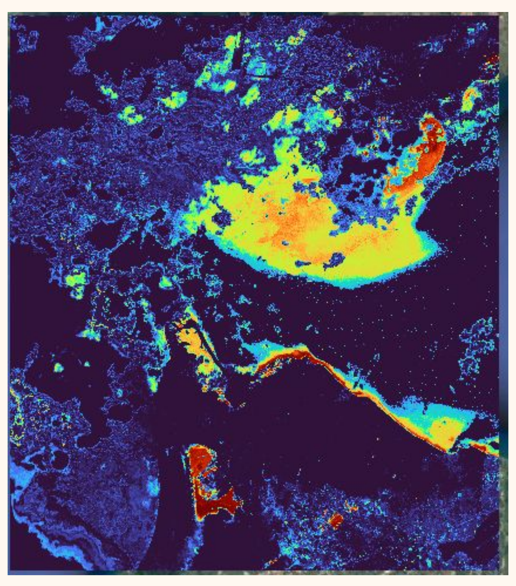
ACs comparison for the muddy water class for 2 scenes (left) & evaluation of annotated pixels (right)

4. Results

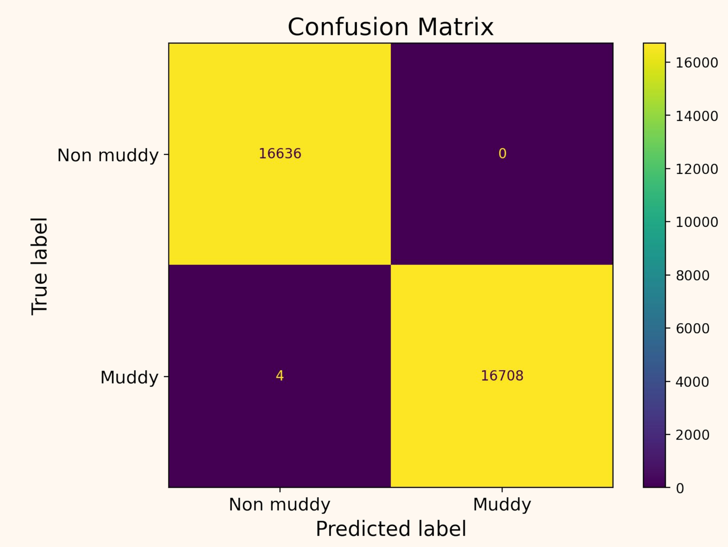
- Classification metrics > 98% (F1-score, Precision, Recall, Accuracy, AUC)
- Good performance in unseen cases
- Overestimation in areas with high chlorophyll
- Highest feature contribution: B11, B12, B4, B8A, B5 (owed to e.g., clouds, vegetation, soil samples)

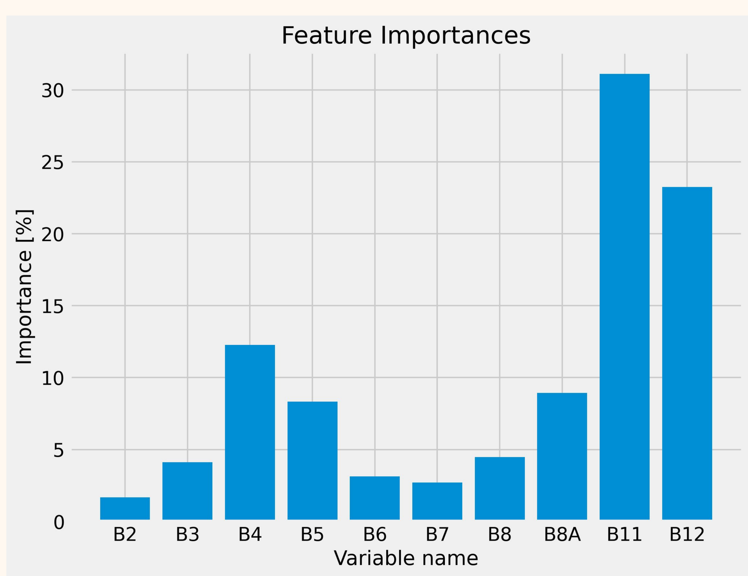






RGB images (left) and prediction probabilities (right)





Confusion matrix (upper) and Random Forest feature importances (lower)

5. Conclusion & Prospect

- Quick muddy water mapping without any calibration is feasible, while room for improvement still exists
- An image-based dataset including global variability will increase model generalization ability
- Comparison of model outputs with water quality parameters (e.g., total suspended sediment)

Presenter: Kostas Vlachos, **kostasvlachosgrs@iti.gr**
Project Coordinator (WQeMS): Ioannis Manakos, **imanakos@iti.gr**

Funded:
This work has been supported by the EU H2020 RIA project WQeMS under grant agreement number 101004157
https://wqems.eu