The role of in-situ measurements to support verification of satellite data products

Presentation of preliminary results

Data Science in Remote Sensing, LTTO.00.027

Group 3

Fjodor Ševtšenko

Fedor Stomakhin

Task definition

- **Task 1**. To compare the reflectance derived from in-situ sensor with satellite derived estimates over vegetation target.
  - Fjodor Ševtšenko (current presentation)
- **Task 2**. To investigate, how well the photodiode signal correlates with Ed values.
  - Fedor Stomakhin (next presentation)

Structure

1. Give clear overview of the topic – why is it important / what is it about?
2. Give clear overview of the data – what data was used?
3. Analyses of the data – what you used, how processed etc.
4. Results.
5. GitHub references.

1. Give clear overview of the topic – why is it important / what is it about?

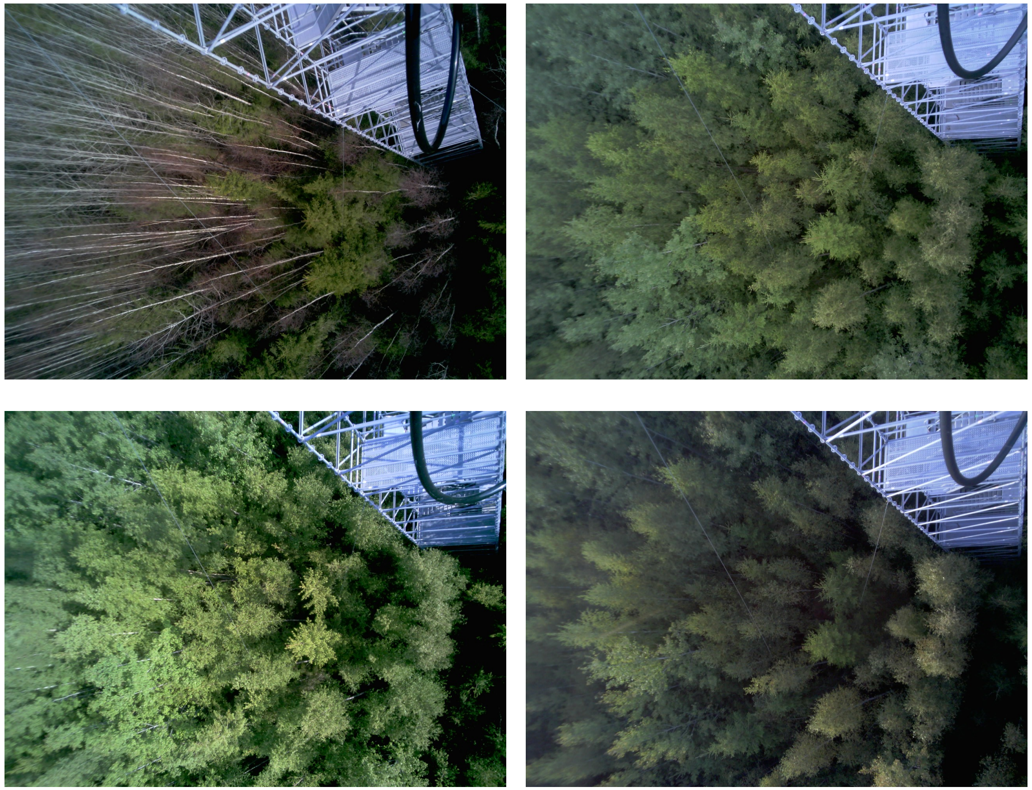
- Mainly about ***integration*** of ground-based measurements with satellite-derived data to enhance the accuracy and reliability of information.
- Importance
  1. Data validation and calibration
     - Satellite products; errors due to atmospheric conditions or sensor degradation.
     - In-situ data help to identify and correct such errors.
  2. Temporal and spatial complementary
     - Satellite data: wide spatial coverage, but limitation in capturing high temporal variability.
     - In-situ data: continues monitoring; complementing the temporal coverage of satellites.
  3. Economical reasons
- Applications
  - Validation of models based on satellite data (e.g. hydrological, ecological)
  - Disaster management
  - Nature resource management (e.g., water, forest)

2. Data overview

- In-situ sensor, forest site Järvselja
  - Location coordinates <- { ‘latitude’: 58,281975, ‘longitude’: 27,312960 }
  - Hyperspectral <- { ‘wavelength’: 350 – 1686 }
  - 2023/05 – 2023/08

- Satellite data
  - ESTHub Processing Platform
  - Bands <- [‘B1’, ‘B2’, ‘B3’, ‘B4’, ‘B5’, ‘B6’, ‘B7’, ‘B8’, ‘B8A’, ‘B9’, ‘B11’]
  - 2023/05 – 2023/08
  - Request <-
    - ‘satellite’: ‘Sentinel-2 MSI L2’
    - ‘latitude’: 58,281975, ‘longitude’: 27,312960
    - ‘macro pixel size’: 5
    - ‘maximum time difference’: 0
    - ‘filtered mean coefficient’: 1.5
    - ‘grouping column’: ‘SITE’
    - ‘percentage of allowed falling products’: 50

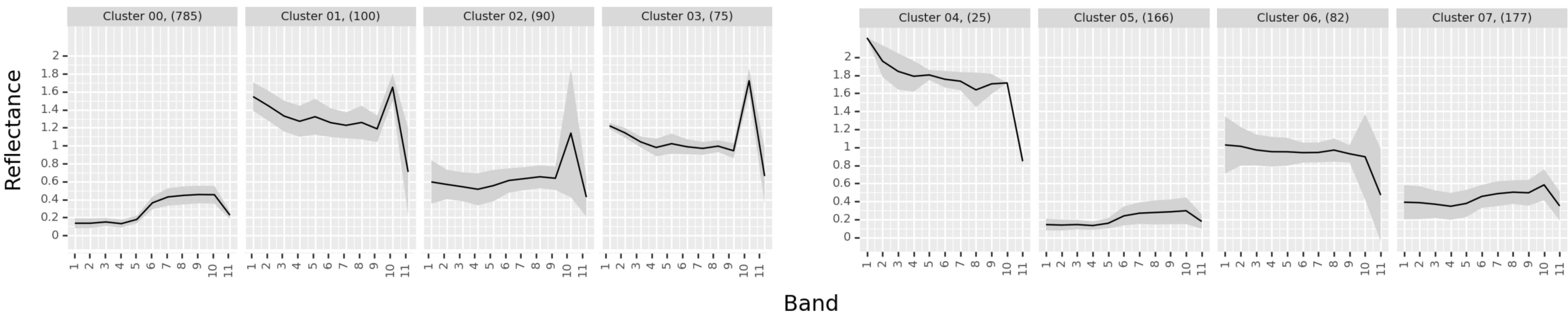
2.1 Data overview: Sentinel2MsiL2, in-situ data, Hypernets



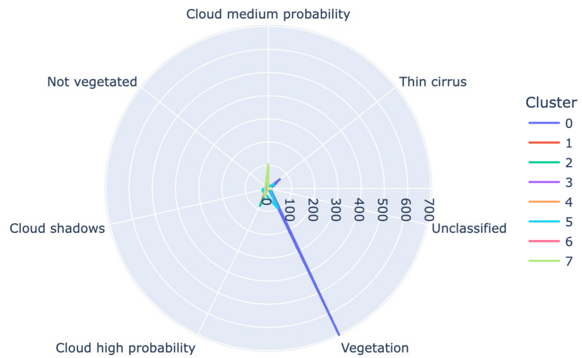
3. Analyses of the data

1. In-situ data
   - NetCDF preprocessing
   - UTC matching intervals calculation (1h, 2h, 3h, 4h)
   - Hyperspectral wavelength to multispectral bands convolution
2. Satellite data
   - Kmeans clustering (data distributions grouping, filtering reflectance signatures)
3. In-situ vs satellite data
   - In time alignment
   - Matchups, Pearson correlation
   - Matchups, reflectance signatures by dates comparison
   - Prediction task: given Satellite multispectral reflectance signature, predict In-situ multispectral reflectance signature

4. Results: Kmeans use to filter satellite data



QualitySceneClassification	Cluster QualitySceneClassificationName	0	1	2	3	4	5	6	7
10.0	Thin cirrus	66.0	NaN	NaN	NaN	NaN	NaN	4.0	34.0
3.0	Cloud shadows	NaN	NaN	NaN	NaN	NaN	NaN	NaN	29.0
4.0	Vegetation	700.0	NaN	NaN	NaN	NaN	NaN	3.0	107.0
5.0	Not vegetated	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN
7.0	Unclassified	9.0	NaN	NaN	NaN	NaN	NaN	20.0	5.0
8.0	Cloud medium probability	1.0	50.0	NaN	2.0	25.0	NaN	106.0	NaN
9.0	Cloud high probability	NaN	32.0	25.0	88.0	75.0	75.0	42.0	NaN

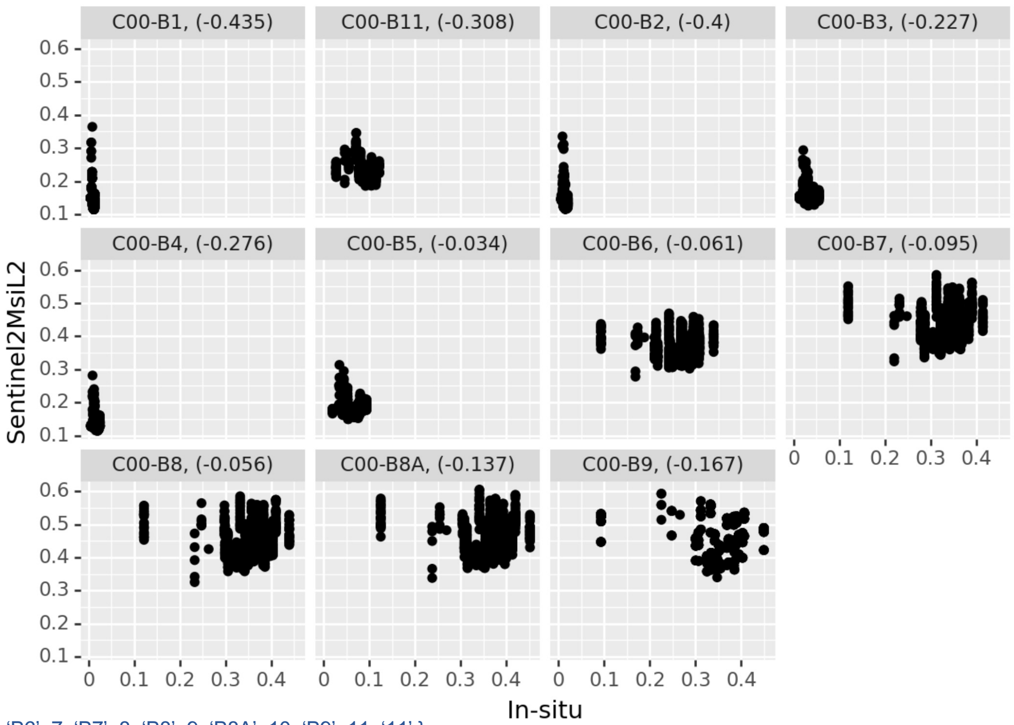


Bands: { 1: ‘B1’, 2: ‘B2’, 3: ‘B3’, 4: ‘B4’, 5: ‘B5’, 6: ‘B6’, 7: ‘B7’, 8: ‘B8’, 9: ‘B8A’, 10: ‘B9’, 11: ‘11’ }

Central wavelength/bandwidth: { ‘B1’: 442.7±20, ‘B2’: 492.7±65, ‘B3’: 559.8±35, ‘B4’: 664.6±30, ‘B5’: 704.1±14, ‘B6’: 740.5±14, ‘B7’: 782.8±14, ‘B8’: 832.8±105, ‘B8A’: 864.7±21, ‘B9’: 945.1±19, ‘B11’: 1613.7± 90 }

Cluster 0, vegetation – in the near-infrared region (approx. 700-1300 nanometers), vegetation reflectance increases significantly (Healthy vegetation reflects a large portion of NIR light due to the strong scattering by cellular structures within leaves and other plant components.).

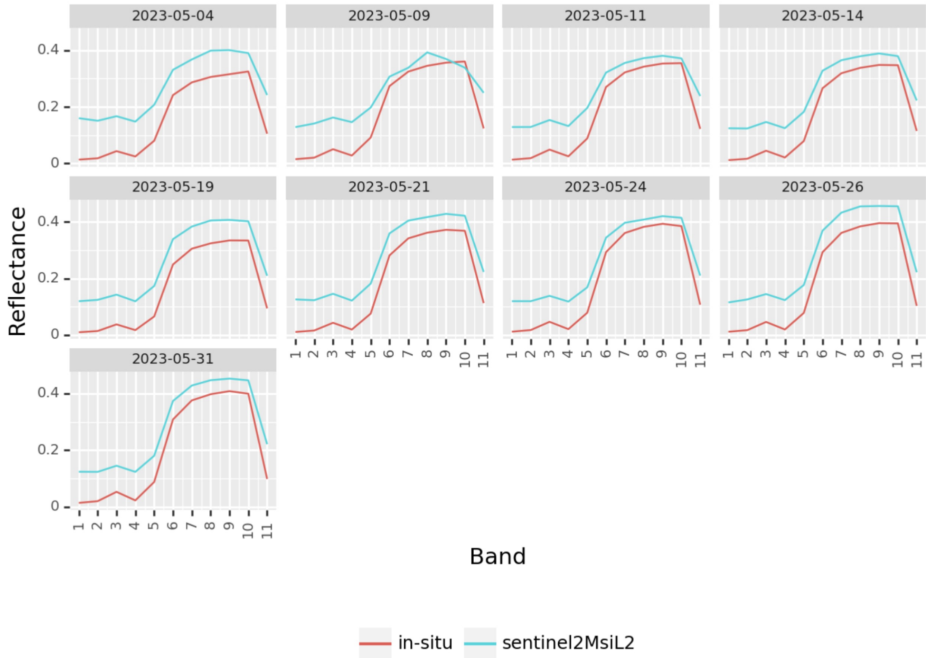
4.1 Results: matchups, Pearson correlation, cluster 0 (vegetation)



Bands: { 1: 'B1', 2: 'B2', 3: 'B3', 4: 'B4', 5: 'B5', 6: 'B6', 7: 'B7', 8: 'B8', 9: 'B8A', 10: 'B9', 11: '11' }

Central wavelength/bandwidth: { 'B1': 442.7±20, 'B2': 492.7±65, 'B3': 559.8±35, 'B4': 664.6±30, 'B5': 704.1±14, 'B6': 740.5±14, 'B7': 782.8±14, 'B8': 832.8±105, 'B8A': 864.7±21, 'B9': 945.1±19, 'B11': 1613.7± 90 }

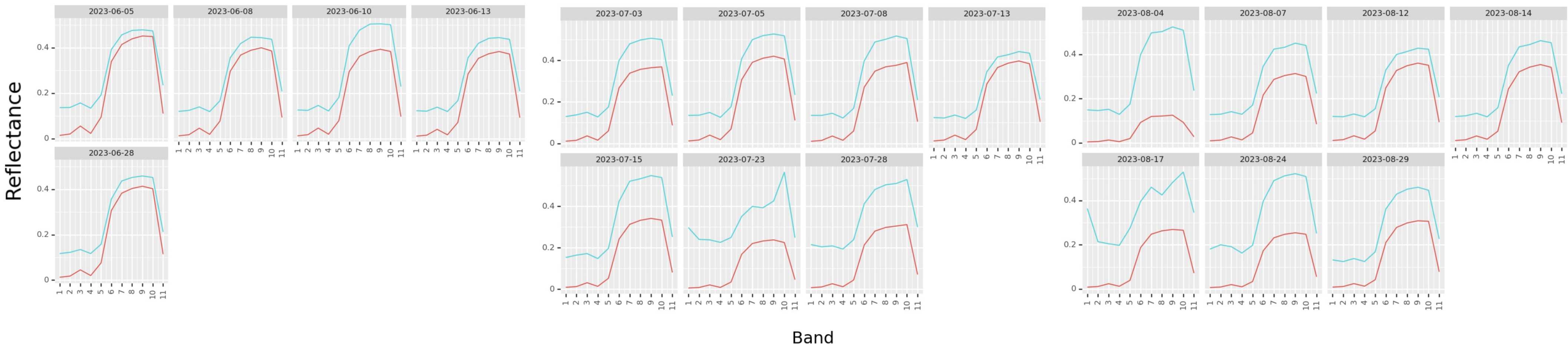
4.2 Results: matchups, reflectance signature by dates, 2023-05, cluster 0 (vegetation)



Bands: { 1: ‘B1’, 2: ‘B2’, 3: ‘B3’, 4: ‘B4’, 5: ‘B5’, 6: ‘B6’, 7: ‘B7’, 8: ‘B8’, 9: ‘B8A’, 10: ‘B9’, 11: ‘11’ }

Central wavelength/bandwidth: { ‘B1’: 442.7±20, ‘B2’: 492.7±65, ‘B3’: 559.8±35, ‘B4’: 664.6±30, ‘B5’: 704.1±14, ‘B6’: 740.5±14, ‘B7’: 782.8±14, ‘B8’: 832.8±105, ‘B8A’: 864.7±21, ‘B9’: 945.1±19, ‘B11’: 1613.7± 90 }

4.3 Results: matchups, reflectance signature by dates, 2023-06/07/08, cluster 0 (vegetation)



Bands: { 1: ‘B1’, 2: ‘B2’, 3: ‘B3’, 4: ‘B4’, 5: ‘B5’, 6: ‘B6’, 7: ‘B7’, 8: ‘B8’, 9: ‘B8A’, 10: ‘B9’, 11: ‘11’ }

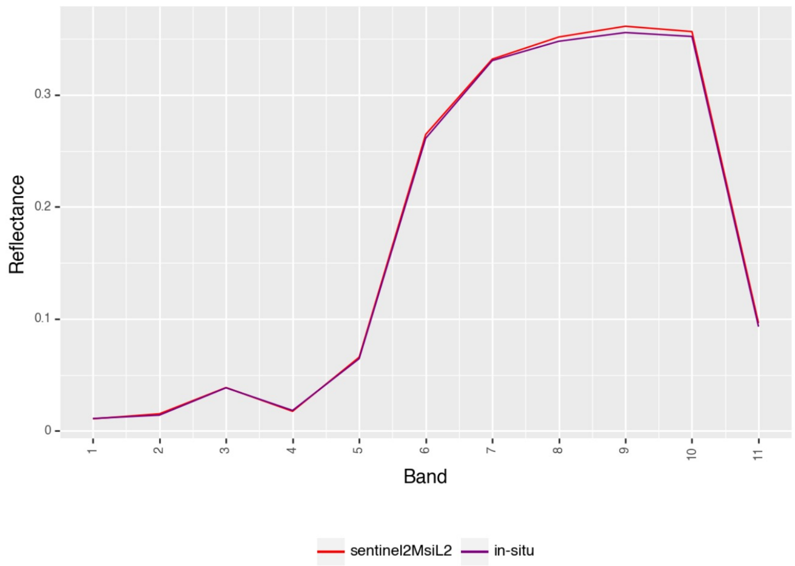
Central wavelength/bandwidth: { ‘B1’: 442.7±20, ‘B2’: 492.7±65, ‘B3’: 559.8±35, ‘B4’: 664.6±30, ‘B5’: 704.1±14, ‘B6’: 740.5±14, ‘B7’: 782.8±14, ‘B8’: 832.8±105, ‘B8A’: 864.7±21, ‘B9’: 945.1±19, ‘B11’: 1613.7± 90 }

4.4 Results: prediction task, cluster 0 (vegetation)

```python
from torch import nn

class MultiOutputNetv1(nn.Module):
    def __init__(self):
        super().__init__()
        self.fc1 = nn.Linear( in_features: 11,  out_features: 50)
        self.relu1 = nn.ReLU()
        self.fc2 = nn.Linear( in_features: 50,  out_features: 100)
        self.relu2 = nn.ReLU()
        self.fc3 = nn.Linear( in_features: 100,  out_features: 50)
        self.relu3 = nn.ReLU()
        self.fc4 = nn.Linear( in_features: 50,  out_features: 11)  # Output layer with 11 neurons

    def forward(self, x):
        x = self.relu1(self.fc1(x))
        x = self.relu2(self.fc2(x))
        x = self.relu3(self.fc3(x))
        x = self.fc4(x)
        return x
```



- Data points - 1025
- Epochs – 100
- Learning rate – 0.001
- Train/test – 80/20%
- MSE, train – 0.0009
- MSE, test – 0.0008
- RMSE, test – 0.0281

Bands: { 1: 'B1', 2: 'B2', 3: 'B3', 4: 'B4', 5: 'B5', 6: 'B6', 7: 'B7', 8: 'B8', 9: 'B8A', 10: 'B9', 11: '11' }

Central wavelength/bandwidth: { 'B1': 442.7±20, 'B2': 492.7±65, 'B3': 559.8±35, 'B4': 664.6±30, 'B5': 704.1±14, 'B6': 740.5±14, 'B7': 782.8±14, 'B8': 832.8±105, 'B8A': 864.7±21, 'B9': 945.1±19, 'B11': 1613.7± 90 }

5. GitHub references

- **hypernesLJaesL2A**
  - /_out00/hypernetsLJaesL2A_out00_001_0.ipynb
  - /_out01/hypernetsLJaesL2A_out01_001_0 - calculate utc intervals.ipynb
  - /_out01/hypernetsLJaesL2A_out01_002_0 - spectral convolution.ipynb

- **sentinel2MsiL2**
  - /_out01/sentinel2MsiL2_out01_001_0 - kmeans.ipynb
  - /_out01/sentinel2MsiL2_out01_002_0 - in time alignment.ipynb
  - /_out01/sentinel2MsiL2_out01_003_0 - matchup.ipynb

Thank you!