AI4EO Activities at CommSensLab-UPC

Adriano Camps, Mercè Vall·llosera, Carlos López-Martínez
and students and collaborators
6 Febrero 2023
Earth Observation @ CommSensLab – UPC

AI4EO Activities and/or Applications

- Soil Moisture Estimation
- Forest Fire Burned Area Prediction
- Dengue Prediction in Brazil
- Multitemporal SAR Coherence for Land Mapping/Classification
- Forest Height Estimation based on SAOCOM L-band SAR Data
- Soil Moisture, Sea Ice Extent, Concentration and Thickness, and Sea Surface Salinity retrievals from FSSCat
- HyperSpectral Imagery Compression
CommSensLab has a large experience in (Microwave) Earth Observation from sensors design and development to data analysis and exploitation in diverse applications.

→ Synthetic Aperture Radar (SAR)
  ▪ Collaboration with MWSE

Multifrequency ground-based SAR Sensor
SMOS activities:
from MIRAS instrument to novel applications...

1993
Sea salinity
Soil moisture: downscaled @ 1 km
Sea Ice Thickness up to ~60 cm
Fire risk index

250 m downscaled soil moisture maps

Credits: Miriam Pablos, BEC

2000

2007

2015

GNSS-R:
instruments dev. and applications

RFI detection & mitigation:
MWR & GNSS-R

NanoSats: test bed of new remote sensors
# Soil Moisture Estimation: Random Forest (Data)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Variable</th>
<th>Periodicity</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel 2</td>
<td>12 bands</td>
<td>5 days</td>
<td>60 m</td>
</tr>
<tr>
<td></td>
<td>10 indices</td>
<td>5 days</td>
<td>60 m</td>
</tr>
<tr>
<td>Ancillary data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA5-land Skin Temp.</td>
<td>daily</td>
<td>9 km</td>
<td></td>
</tr>
<tr>
<td>DEM</td>
<td>static</td>
<td>60 m</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>static</td>
<td>60 m</td>
<td></td>
</tr>
<tr>
<td>Aspect</td>
<td>static</td>
<td>60 m</td>
<td></td>
</tr>
<tr>
<td>Hillshade</td>
<td>static</td>
<td>60 m</td>
<td></td>
</tr>
<tr>
<td>Month</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Variable</th>
<th>Variable</th>
<th>Periodicity</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate Change Initiative (CCI)</td>
<td>Soil Moisture (SM)</td>
<td>daily</td>
<td>25 km</td>
</tr>
</tbody>
</table>

- **Study period:** 2018-2019
- **Study area:** central part of the Iberian Peninsula
- **Sentinel 2 (A/B) data** is the most temporally constraining
- **Clouds** can mask Sentinel 2 data
- **ESA CCI** provides a combined SM product (active + passive)
- **ESA CCI SM** has the lowest resolution

---

Soil Moisture Estimation: Random Forest (Methodology & Results)

- **Predictors**
- Aggregation to the target resolution (25 km)
- **Random Forest**
  - 80% training
  - 20% testing
- **Estimation of the 60-m SM maps**

- $R^2 = 0.82$ and RMSE = 0.026 m$^3$ m$^{-3}$
- High spatial heterogeneity
- Difficulty to detect extremes
The location of burned areas was determined using the MODIS burned area dataset.
### Method:

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Resolution</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEC SMOS L3</td>
<td>SM</td>
<td>25 km Ext</td>
<td>Extent of the study by Chaparro et al. (2016), who used remotely observed SM and LST to predict the fires extent</td>
</tr>
<tr>
<td></td>
<td>VOD</td>
<td>25 km Ext</td>
<td></td>
</tr>
<tr>
<td>ERA5 ECMWF</td>
<td>VPD</td>
<td>0.25° Ext</td>
<td>Indicate the aridity conditions in the surface air</td>
</tr>
<tr>
<td>(<a href="https://cds.climate.copernicus.eu/">https://cds.climate.copernicus.eu/</a>)</td>
<td>LST</td>
<td>0.25° Ext</td>
<td></td>
</tr>
<tr>
<td>NCEP–DOE II</td>
<td>$u_{300}$ and $v_{300}$</td>
<td>2.5° Ext</td>
<td>To determine jet stream characteristics in relation to a very big fire using Spatiotemporal Composite technique. Further insight can be found in Jain &amp; Flannigan (2021)</td>
</tr>
<tr>
<td>(<a href="https://psl.noaa.gov">https://psl.noaa.gov</a>)</td>
<td>$\Delta Z_{500}$</td>
<td>2.5° Ext</td>
<td></td>
</tr>
<tr>
<td>SRTM (<a href="https://portal.opentopography.org/">https://portal.opentopography.org/</a>)</td>
<td>Elevation</td>
<td>90 m Ext</td>
<td>Obtained through the United States Geological Survey (USGS)</td>
</tr>
<tr>
<td>GLC2000-JRC (<a href="https://forobs.jrc.ec.europa.eu/">https://forobs.jrc.ec.europa.eu/</a>)</td>
<td>Land use</td>
<td>1 km Ext</td>
<td>Global Land Cover Product (GLC) coordinated by Forest Resources and Carbon Emissions (IFORCE)</td>
</tr>
<tr>
<td>Open Street Map (<a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a>)</td>
<td>Distance to road</td>
<td>- Ext</td>
<td>Calculated in GIS by Euclidean Distance</td>
</tr>
<tr>
<td>MODIS Land Product (<a href="https://lpdaac.usgs.gov/">https://lpdaac.usgs.gov/</a>)</td>
<td>Burned Area</td>
<td>1 km Ext</td>
<td>MODIS burned area datasets (MC64A1) is obtained through sftp (Server: fuoco.geog.umd.edu, Login name: fire, Password: burnt)</td>
</tr>
</tbody>
</table>

**Database:**
- BEC SMOS L3
- ERA5 ECMWF
- NCEP–DOE II
- SRTM
- GLC2000-JRC
- Open Street Map
- MODIS Land Product

**Steps:**
1. Resampling (join Spatial)
2. Burned Area Model (Extra Trees)
3. Isolation Forest (Increasing the outliers by a factor of 2)
4. Burned Area Prediction

Forest Fire Burned Area Prediction (Methodology)

Random Forest (RF) and Extremely randomized Trees (Extra Trees) are both methods that use multiple decision trees to make predictions.

<table>
<thead>
<tr>
<th>Capability</th>
<th>Extra Trees</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of randomness</td>
<td>High, as it uses random thresholds for each feature</td>
<td>Moderate, as just the best split point is used</td>
</tr>
<tr>
<td>Decision tree constructed</td>
<td>All samples are considered for each split</td>
<td>By bootstrap aggregating the samples (that is with replacement)</td>
</tr>
<tr>
<td>Features that are considered for each split</td>
<td>All features</td>
<td>A random subset of features</td>
</tr>
<tr>
<td>Handling noisy and extreme data</td>
<td>Better</td>
<td>Good</td>
</tr>
</tbody>
</table>

**Conclusion:** Extra Trees can be useful when the data is more extreme or contains more noise because the higher level of randomness in the construction of the decision trees.

Illustration of an RF algorithm structure obtained from Serra, (2021)

Forest Fire Burned Area Prediction (Methodology)

Isolation Forest

It uses the concept of decision tree to identify the anomalies (or extreme values). The method is based on the assumption that anomalies are less dense and isolated compared to the normal data points, so the fewer splits required to isolate an instance, the more likely it is to be an anomaly.

Illustration of an Isolation Forest

Outlier: easy to isolate

Regular data point: difficult to isolate
Results: Burned Area Prediction

Forest Fire Burned Area Prediction (Results)

- $R^2 = 0.617$
- $R^2 = 0.675$
- $R^2 = 0.694$
- $R^2 = 0.762$
We separated the fire class categories into low, medium, large, and very large for model comparison and then calculated the evaluation index for all models.

<table>
<thead>
<tr>
<th>Best Model</th>
<th>Evaluation Index</th>
<th>Low fire (≤500 ha)</th>
<th>Medium fire (&gt;500 - 1000 ha)</th>
<th>Large fire (&gt;1000 - 3000ha)</th>
<th>Very large fire (&gt;3000 ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Trees + Isolation Forest</td>
<td>Accuracy</td>
<td>0.86</td>
<td>0.61</td>
<td>0.46</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.79</td>
<td>0.63</td>
<td>0.52</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.86</td>
<td>0.61</td>
<td>0.46</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>F1-Score</td>
<td>0.82</td>
<td>0.62</td>
<td>0.49</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Dengue Prediction in Brazil

- **Environmental data from satellite**: NDVI, NDWI and LST, accumulated precipitation (in mm), Soil moisture
- **Dengue episodes distribution data**: Notifiable Diseases Information System (SINAN), developed by Ministry of Health of Brazil and available at DATASUS
- The actual risk

**NDVI**                         accumulated precipitation

**Urban typology classification**

**Municipal Human Development Index**

**10 most decisive parameters**

<table>
<thead>
<tr>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Risk</td>
</tr>
<tr>
<td>Precipitation</td>
</tr>
<tr>
<td>Day Temperature</td>
</tr>
<tr>
<td>Night Temperature</td>
</tr>
<tr>
<td>NDWI</td>
</tr>
<tr>
<td>Soil Moisture</td>
</tr>
<tr>
<td>Federal state</td>
</tr>
<tr>
<td>NDVI</td>
</tr>
<tr>
<td>IDHM</td>
</tr>
<tr>
<td>Urban typology</td>
</tr>
</tbody>
</table>

**Risk indexes maps obtained for March 2013 using the registered data (left) and the random forest model (right)**

**INDEX A | VALUES Cases per 100000 inhab**
---|---
Minimum | < 100
Low | >100, <200
Medium | >200, <300
High | >300, <400
Very High | >400
Multitemporal SAR Coherence for Land Mapping/Classification

SinCohMap Project: Develop, Analyse and Validate Novel Methodologies for Land Cover & Vegetation Mapping/Classification Using Sentinel-1 Interferometric Coherence Evolution
Multitemporal SAR Coherence for Land Mapping/Classification

Doñana (Spain)
Wetlands & Agriculture
Asc. 2015/2016
51 acquisitions
1275 int. VV, VH

South Tyrol (Italy)
Agriculture, Forest, Mont. region
Asc. 2015/2016
38 acquisitions
730 int. VV, VH

West Wielkopolska (Poland)
Agriculture, Forest
Asc. 2015/2016
56 acquisitions
1540 int. VV, VH

Pre-analysis and training
2015
2016
Min. 12 days temporal baseline

Post-analysis and evaluation
2017
2018
Min. 6 days temporal baseline
Multitemporal SAR Coherence for Land Mapping/Classification

Multi-temporal Coherence Evolution

South Tyrol (Italy)

30/03/2015-13/12/2016
Multitemporal SAR Coherence for Land Mapping/Classification

Round-Robin Experience

Pre-processed InSAR stack sites

West Wielkopolska (Poland)  South Tyrol (Italy)  Doñana (Spain)

Land cover maps
### Classification Methodologies Comparison: Classical & ML/DL

- Role of the methodology and role of the input features

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Space Object</th>
<th>Decision Type</th>
<th>Temporal Baselines</th>
<th>Polarization</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Forest</strong></td>
<td>Pixel</td>
<td>ML Classifier</td>
<td>shortest</td>
<td>VV &amp; VH</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Eigen-value Decomposition + RF</strong></td>
<td>Pixel</td>
<td>ML Classifier</td>
<td>all</td>
<td>VV &amp; VH</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Temporal Dynamic Indices + RF</strong></td>
<td>Pixel</td>
<td>ML Classifier</td>
<td>all</td>
<td>VV &amp; VH</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Object-based (KTH-SEG) SVM</strong></td>
<td>Object</td>
<td>ML Classifier</td>
<td>two shortest</td>
<td>VV &amp; VH</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Super-pixel (SLIC) + kNN</strong></td>
<td>Object</td>
<td>ML Classifier</td>
<td>all</td>
<td>VV &amp; VH</td>
<td>No</td>
</tr>
<tr>
<td><strong>Expert Knowledge Decision Tree</strong></td>
<td>Object</td>
<td>Decision Tree</td>
<td>shortest</td>
<td>VV &amp; VH</td>
<td>No</td>
</tr>
<tr>
<td><strong>Data Adaptive Rule-Based</strong></td>
<td>Object</td>
<td>Threshold</td>
<td>selection</td>
<td>VV &amp; VH</td>
<td>No</td>
</tr>
</tbody>
</table>
### Impact of Polarimetric & Interferometric SAR Information

<table>
<thead>
<tr>
<th>Location</th>
<th>Coherence</th>
<th>Intensity</th>
<th>Coherence + Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Doñana</strong> (Spain)</td>
<td>78.2</td>
<td>74.6</td>
<td>81.9</td>
</tr>
<tr>
<td></td>
<td>74.5</td>
<td>74.0</td>
<td>79.8</td>
</tr>
<tr>
<td></td>
<td>79.8</td>
<td>77.8</td>
<td>83.3</td>
</tr>
<tr>
<td><strong>South Tyrol</strong> (Italy)</td>
<td>70.9</td>
<td>54.9</td>
<td>72.0</td>
</tr>
<tr>
<td></td>
<td>68.8</td>
<td>56.1</td>
<td>70.9</td>
</tr>
<tr>
<td></td>
<td>72.5</td>
<td>58.9</td>
<td>73.8</td>
</tr>
<tr>
<td><strong>West Wielkopolska</strong> (Poland)</td>
<td>71.0</td>
<td>64.0</td>
<td>73.2</td>
</tr>
<tr>
<td></td>
<td>67.4</td>
<td>66.9</td>
<td>71.7</td>
</tr>
<tr>
<td></td>
<td>71.7</td>
<td>69.5</td>
<td>75.0</td>
</tr>
</tbody>
</table>

**Methodology:** Eigen-decomposition + Random Forest
Analysis of Temporal Features in a Random Forest Approach

Global feature importance: Importance of each feature in the final classification accuracy

Lack of information about how much each feature contributes to each class classification accuracy and lack of temporal information
Feature importance per class: Importance of each feature in the final classification accuracy
Multi-temporal SAR Coherence for Land Mapping/Classification

Temporal features are selected specifically:

- Yearly
- Minimum temporal samples 10 & 20 in identified times detected as important
- Minimum temporal samples 10 & 20 in random times

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of estimators</th>
<th>Max. num. features</th>
<th>Number of Images</th>
<th>Number of features</th>
<th>Overall Accuracy</th>
<th>Macro Average</th>
<th>Weighted Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>100</td>
<td>23</td>
<td>49</td>
<td>519</td>
<td>88.99%</td>
<td>75.86%</td>
<td>87.27%</td>
</tr>
<tr>
<td>➔ Optimum 2015*</td>
<td>400</td>
<td>50</td>
<td>49</td>
<td>519</td>
<td>89.89%</td>
<td>78.02%</td>
<td>88.44%</td>
</tr>
<tr>
<td>2016*</td>
<td>400</td>
<td>50</td>
<td>29</td>
<td>319</td>
<td>86.10%</td>
<td>67.31%</td>
<td>83.80%</td>
</tr>
<tr>
<td>➔ 10**</td>
<td>400</td>
<td>50</td>
<td>10</td>
<td>70</td>
<td>87.06%</td>
<td>68.89%</td>
<td>85.29%</td>
</tr>
<tr>
<td>20**</td>
<td>400</td>
<td>50</td>
<td>20</td>
<td>138</td>
<td>88.75%</td>
<td>74.77%</td>
<td>87.11%</td>
</tr>
<tr>
<td>10 random***</td>
<td>400</td>
<td>50</td>
<td>10</td>
<td>66</td>
<td>82.69%</td>
<td>57.03%</td>
<td>80.05%</td>
</tr>
<tr>
<td>20 random***</td>
<td>400</td>
<td>50</td>
<td>20</td>
<td>136</td>
<td>86.98%</td>
<td>69.89%</td>
<td>84.89%</td>
</tr>
</tbody>
</table>

* Year representation  **Set of relevant temporal samples  ***Set of random temporal samples

Temporal samples reduced by 80%
but classification accuracy only drops 3%

Reduction of computation power & storage
Forest Height Estimation based on SAOCOM L-band SAR Data

Study test sites with Multitemporal L-band SAR data, GEDI data and Ground-Truth

Area 1: Corrientes

Area 2: Neuquén
Interferometric baseline & system effects (C vs L-band)

Saocom: 16 días
Sentinel-1: 12 días

Saocom: 48 días
Sentinel-1: 48 días
Coherence vs. GEDI Forest Height

Canopy height / HoA vs. coherence
Corrientes and Neuquen

8-day SAOCOM Coherence
First inversion results

Campo Aurora Celeste (BDP) - Corrientes

Pinus Taeda - 18 años
FSSCat Mission (i): SM/SIE/SIC/SIT/SSS retrievals

- **FSSCat mission:**
  - 3\textsuperscript{Cat}-5/A: MWR + GNSS-R (GPS+Gal) + RFI detection/mitigation
  - 3\textsuperscript{Cat}-5/B: HyperScout-2 (VNIR+TIR Hyperspectral Imager) + AI Proc
  - Both: O-ISL + RF-ISL

[https://youtu.be/lQAaoYUPluA](https://youtu.be/lQAaoYUPluA)
FSSCat Mission (ii): SM/SIE/SIC/SIT/SSS retrievals

- FMPL-2 onboard 3Cat-5/A simultaneously collected GNSS-R and L-band radiometry data to retrieve:
  - Soil Moisture (4 x ANNs): Optical data only, Optical + L-band MWR data (as in SMOS), GNSS-R data (e.g. NASA CyGNSS), and GNSS-R + L-band MWR
  - Sea Ice Concentration and Extent (2 x ANNs)
  - Sea Ice Thickness (1 x ANN): most difficult to train! Probably because high non-linearities
  - Wind Speed (1 x ANN) and Sea Surface Salinity (1 x ANN)

- All perform well, providing scientific quality data, with a modest budget mission

- ANNs can be applied to retrieve these and other geophysical variables from GNSS-R data where it is difficult to capture the GMF (e.g. vegetation height [1]). However, in some cases they do not outperform classical analysis (i.e. “Natural Intelligence”, e.g. snow thickness and sea ice thickness in MOSAIC [2], or altimetry [3]).

HyperSpectral Imagery Compression by Sequential Band Selection

GOAL: to reduce the down-link requirements in small sats

Indian Pines Ground Truth

Classification using 18 bands out of 220

Error map using 18 bands

$S_i = H_i \cdot \prod_{k}^{k} (1 - \rho_{ij})^w$,

Looks for maximum entropy and inter-band correlation

Current work: do it automatically using NNs without image georegistration