Copernicus Global Land Mapping, from private to public cloud

Bruno Smets, Marcel Buchhorn, Dirk Daems
bruno.smets@vito.be
Outline

• Who are we
• VITO’s mandate in Copernicus
• Global Land Cover mapping in private cloud
• Scaling Up to public cloud
• Conclusions
VITO Remote Sensing

Our MISSION

We make you **see the bigger picture.**
We take you **high-tech with low risk.**
We leverage your **impact on sustainability.**

Our VISION

We go **end-to-end to give you the insights you need.** Without the hassle.
We do your **science & research** to stripe off complexity and risk.
We deliver top-notch **operational services and customer oriented solutions.**
VITO remote sensing

Platforms
- UAV
- AIRBORN
- HALE UAV
- SATELLITE

Sensors

Value Added Services & Information Products

Markets
- Agriculture
- Landuse & Biodiversity
- Climate
- Water & Coast
- Infrastructure
- Security

~90 FTE
~16 M€/yr
VITO Remote Sensing Data Center

Tiered storage capacity:
• ~7 PB Disk storage
• ~5 PB Tape archive

Server capacity:
• ~650 physical servers (2600 core cluster)
• ~300 virtual servers

Networking capacity:
• Géant connection of 10 GiB

Technologies used:
- openstack
- ceph
- jupyter
- docker
- vmware
- GeoServer
- FOREMAN

Customers:
- European Commission
- ESA
- FAO
- United Nations
VITO MANDATE IN COPERNICUS
VITO in Copernicus


• Lead Climate (C3S) Land Biosphere ECV (since 2017)

• Member of Copernicus Academy
• Member of Copernicus Relays network
• Collaborative Ground Segment for Sentinels ([http://www.terrascope.be](http://www.terrascope.be))
CGLOPS Product Portfolio

**VEGETATION**
- Leaf Area Index (LAI)
- Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)
- Fraction of vegetation cover (FCOVER)
- Normalized Difference Vegetation Index (NDVI)
- Vegetation Condition Index
- Vegetation Productivity Index
- Dry Matter Productivity
- Burnt Area
- Greenness Evolution Index
- Phenology metrics
- Moderate Yearly Land Cover

**ENERGY**
- Top-of-Canopy reflectance
- Surface Albedo
- Land Surface Temperature
- Radiation Fluxes
- Evapotranspiration
- Active Fires
- Surface soil moisture
- Soil Water Index

**WATER**
- Water Bodies
  - Coastal Erosion
  - Lake surface water temperature
  - Lake and river water level*
  - Lake surface reflectance
  - Lake turbidity
  - Lake trophic state

**CRYOSPHERE**
- Lake Ice Extent
- Snow cover extent
- Snow water equivalent

* non-gridded product
Dynamic Global Land Cover Map

A systematic SERVICE providing a DYNAMIC, YEARLY, USER-ORIENTED at GLOBAL scale @ 100m resolution from 2015 onwards

Available collections

<table>
<thead>
<tr>
<th>Algorithm version</th>
<th>Spatial</th>
<th>Temporal</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Africa</td>
<td>2015</td>
<td>PROBA-V</td>
</tr>
<tr>
<td>2*</td>
<td>Global</td>
<td>2015 - present</td>
<td>PROBA-V, Sentinel-2</td>
</tr>
</tbody>
</table>

(*) under preparation, expected release 2019

Contributors

- Marcel BUCHHORN, Luc BERTELS, Ruben VAN DE KERCHOVE, Bruno SMETS
- Martin HEROLD, Nandika TSENDBAZAR, Jan VERBESSELT, Dainius MASILIUNAS
- Steffen FRITZ, Myroslava LESIV, Martina DUERAUER
» Pre-processing
  » Data cleaning & fusion
  » Generation of TS metrics
» Classifier/Regressor
  » RF optimized, 5 folded CV
  » 400 metrics, best band selection per 'eco'-zone
  » 25K training points
  » 5 results + 4 fraction maps
» Expert Rules
  » Decision tree
  » Incorporate areas of agreement from GLC maps
  » Imprint JRC GSW/GHSL, DLR GSL (GUF+)
Flexible map - tailored to your application need
https://blog.vito.be/remotesensing/lcm-tailored-to-your-needs

- No ‘fixed’ mixed classes
- Continuous cover fractions

CEOS-LPV ‘independent’ validated, incl. spatial distributions
- Overall accuracy: 72.3% +/- 1.8% to 80.4% +/- 1.1%
- MAE and RMSE: 6% to 14% on cover fractions
COPERNICUS LAND COVER MAPPING IN PRIVATE CLOUD
Copernicus LC100

- For V1 of the Africa product
  - Fully automated workflow, with ~20K training points over Africa continent

- For V2 of the Global product -> less geometric distortion & better interoperability – Landsat and Sentinel-2
  - Generation of PROBA-V UTM 100m ARD
  - Training data and validation data was also switched to UTM (120K human qualitative training points with 10x10 box of 10m resolution)

- For V2 we integrate change detection
  - BFAST method for detecting and characterizing change within time series
  - NAUC indicator on stability of pixels
PROBA-V to UTM

- geometric correction including pixel alignment to the Sentinel-2 grid
- cloud and cloud shadow detection PLUS snow/ice detection (using additional ancillary data) – NEW: *temporal cloud detection*
- Improved atmospheric correction using CAMS NRT data for AOD550, O3, and TCWV
- image compositing to generate time series stacks plus *image clipping* to the Sentinel-2 tiling grid (resolving issues in overlapping areas)
- generation of final status masks – improved to usable data mask
- **gap-filling** in the time series of a pixel via a Kalman-filter approach which uses the PROBA-V 300m corresponding side cameras information, compositing to 5-daily observations
A node in a federation of platforms
PROBA-V to UTM

- Over **50 million** single acquisitions were processed to generate a global PROBA-V ARD at 100m resolution spanning over 5 years ... in 2 weeks using less 30% of resources
- Fully **aligned** to the **Sentinel-2** tiling grid and naming
- Additional metadata are injected to enable a SpatioTemporal Asset Catalog (STAC), based on an ElasticSearch database to allow easy search through json
- Geotrellis based backend, allowing on the fly extraction of time-series cubes for specified areas or whole tiles – used in OpenEO
LC100 Classification

- Implemented in **15 python** modules
- 3 years data in ‘base’ mode
- **270 metrics** per tile
- 18962 tiles:
  - 5 folded CV
  - 6 classifiers + 7 regressors (+urban, water, crop)
- Up till now **31 continental** runs performed
- Using 500 executors, within 1GB memory
  - Data preparation takes 2-3 days
  - Classification < 1 day
SCALING UP TO PUBLIC CLOUD
Continuity & Higher spatial resolution
Sensor agnostic (fusion), re-use LC workflow

- **PROBA-V**
  - L1C 100m
  - L3 ARD 100m
- **PROBA-V**
  - L1C 300m
  - L3 ARD 300m
- **Sentinel-2**
  - L1C 10/20m
  - L2B ARD
- **Sentinel-1**
  - SLC / GRD 20m
  - L1C F0, Coh ARD
- **LANDSAT**
  - L1C 30m
  - L2B ARD

**Training**
- GeoWIKI Training Data

**Ancillary**
- Diverse Ancillary Datasets

**LC pre-processing**
- **Clean** (madHANTS)
- **Composite** (5d, 10d)
- **Data Fusion** (Kalman)
- **Metrics ARD**
  - Spectral
  - Texture
  - Phenologic
  - Statistical

**LC classification & post-processing**
- Training data preparation
- Metrics Selection (per cluster)
- RF Classifier & regressor
- Expert Rules
- Ancillary data preparation
CLOSING
Where are we?

- Full automated Land Cover workflow
  - User-customization in classes possible through 4->7 covers
  - Include Spatial accuracy maps (& Change maps)
  - Large 10m database of high quality training (&validation) points
  - Sensor agnostic
    - Global on PROBA-V UTM
    - AOI on Sentinel & Landsat

- Highly optimized workflow in Spark
  - Stable in private cloud environment
  - Infrastructure agnostic (Spark standalone)
  - Tests in public cloud environment ongoing
Next steps

- **LC100 v2** (Global 2015 + Africa change)
  - Release at Living Planet Symposium
  - Sneak preview V2
    - + 7 Cover Fractions
    - + Data Density Indicator
    - + Spatial Accuracy Map

- **LC100 v3**
  - Global Sentinel (1+2)
  - Explore/integrate new AI/ML techniques (reCNN)
THANK YOU

remotesensing.vito.be