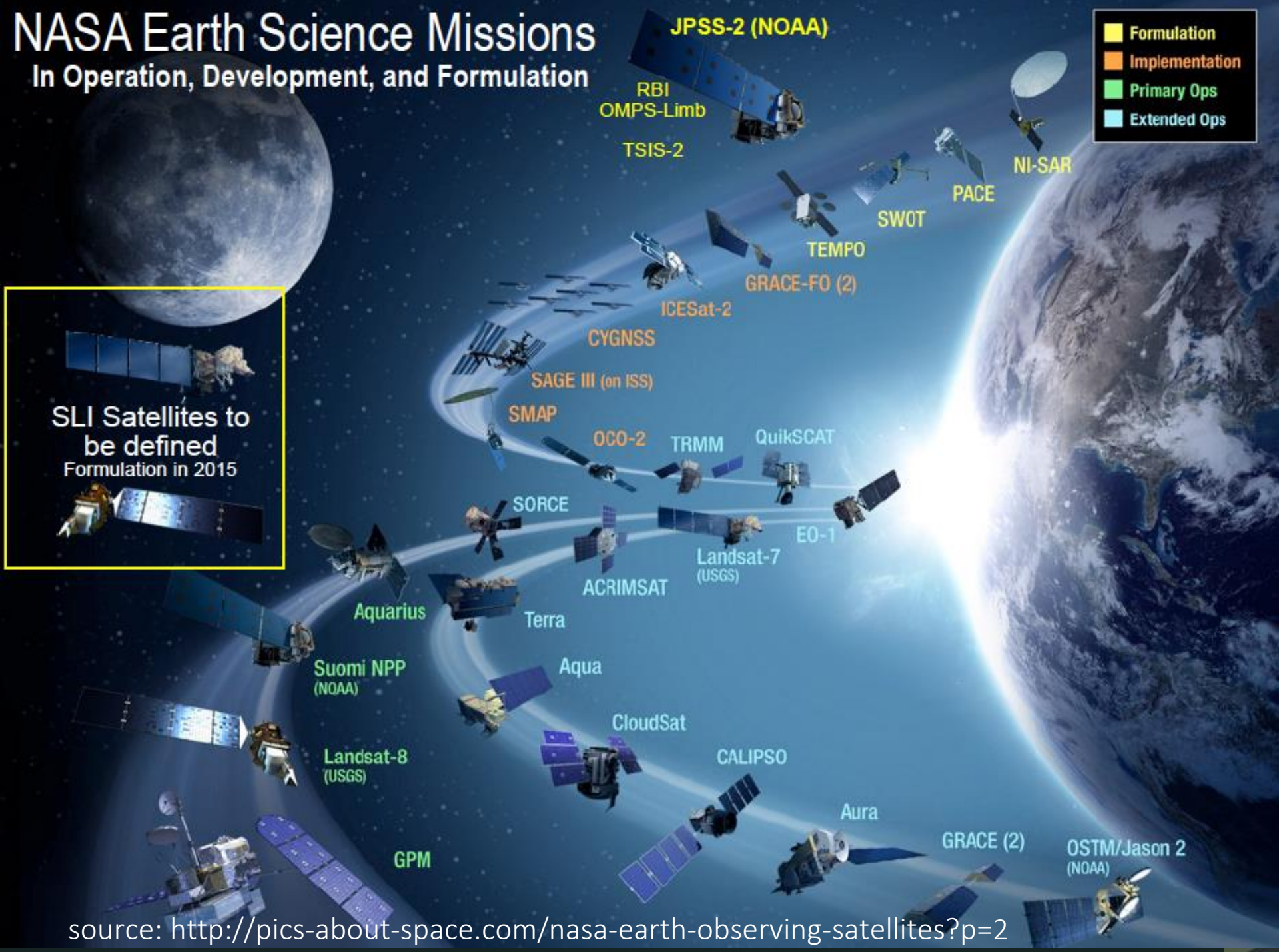
Satellite image processing in a HPC environment

Markus Müller, James Shepherd, Sam Gillingham

Landcare Research
Manaaki Whenua

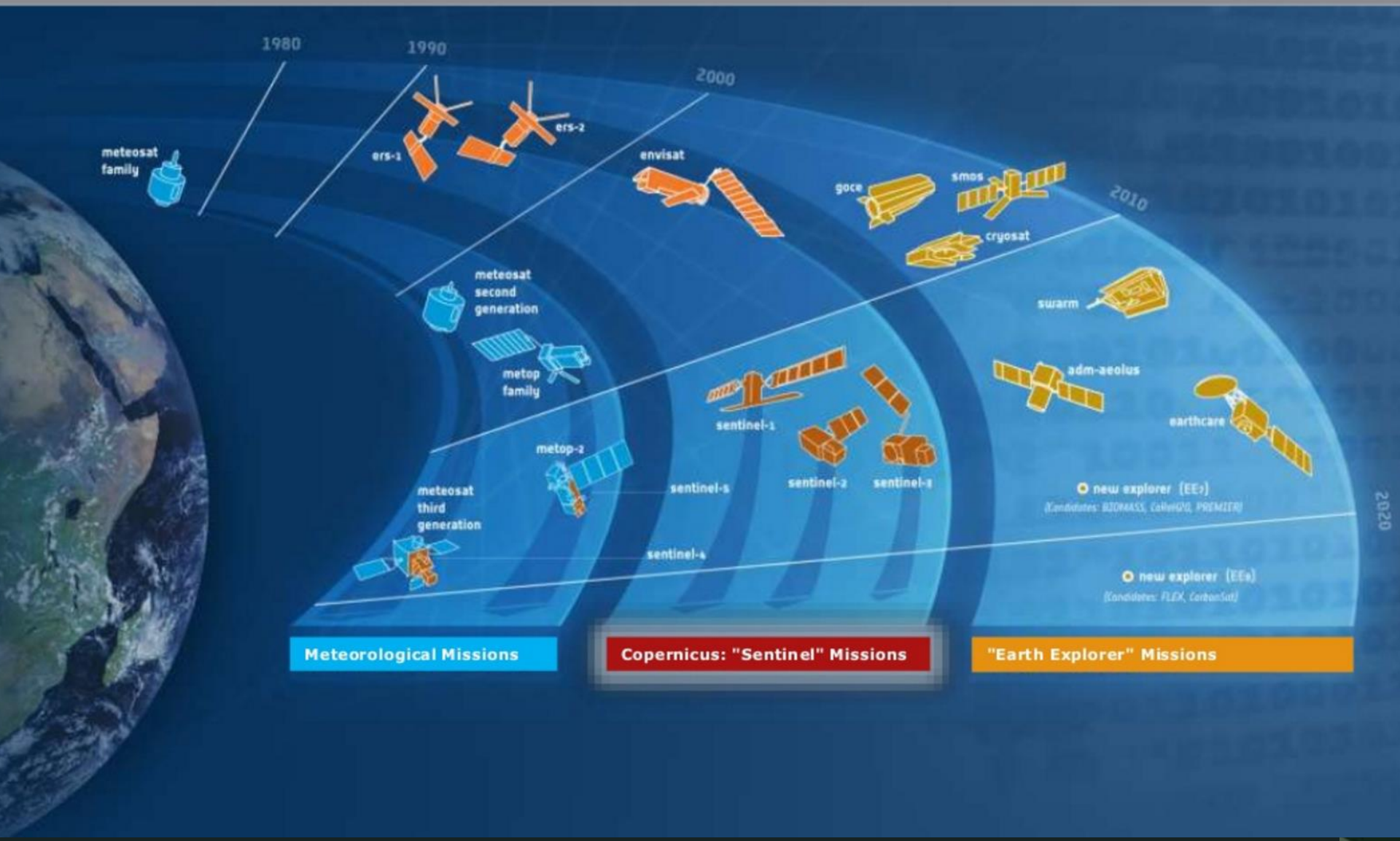
Agenda

- The data explosion in earth observation
- Short intro to satellite remote sensing
- Examples implemented on Pan cluster
  - Pre-processing of Landsat images
  - MODIS land surface temperature analysis
  - Woody patches / LCDB improvement
- Conclusions



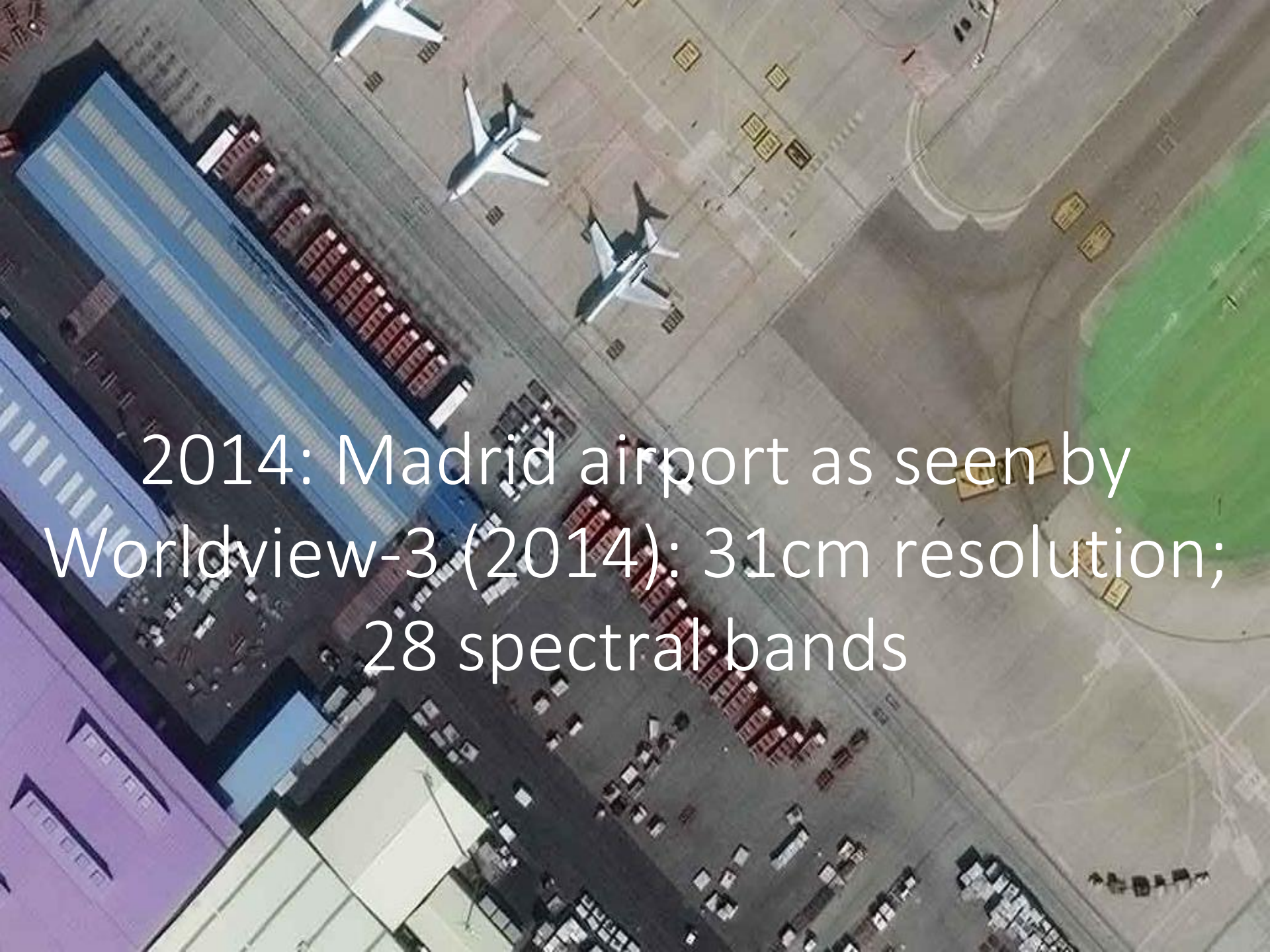
source: http://pics-about-space.com/nasa-earth-observing-satellites?p=2

The ESA Earth Observation Satellite Fleet



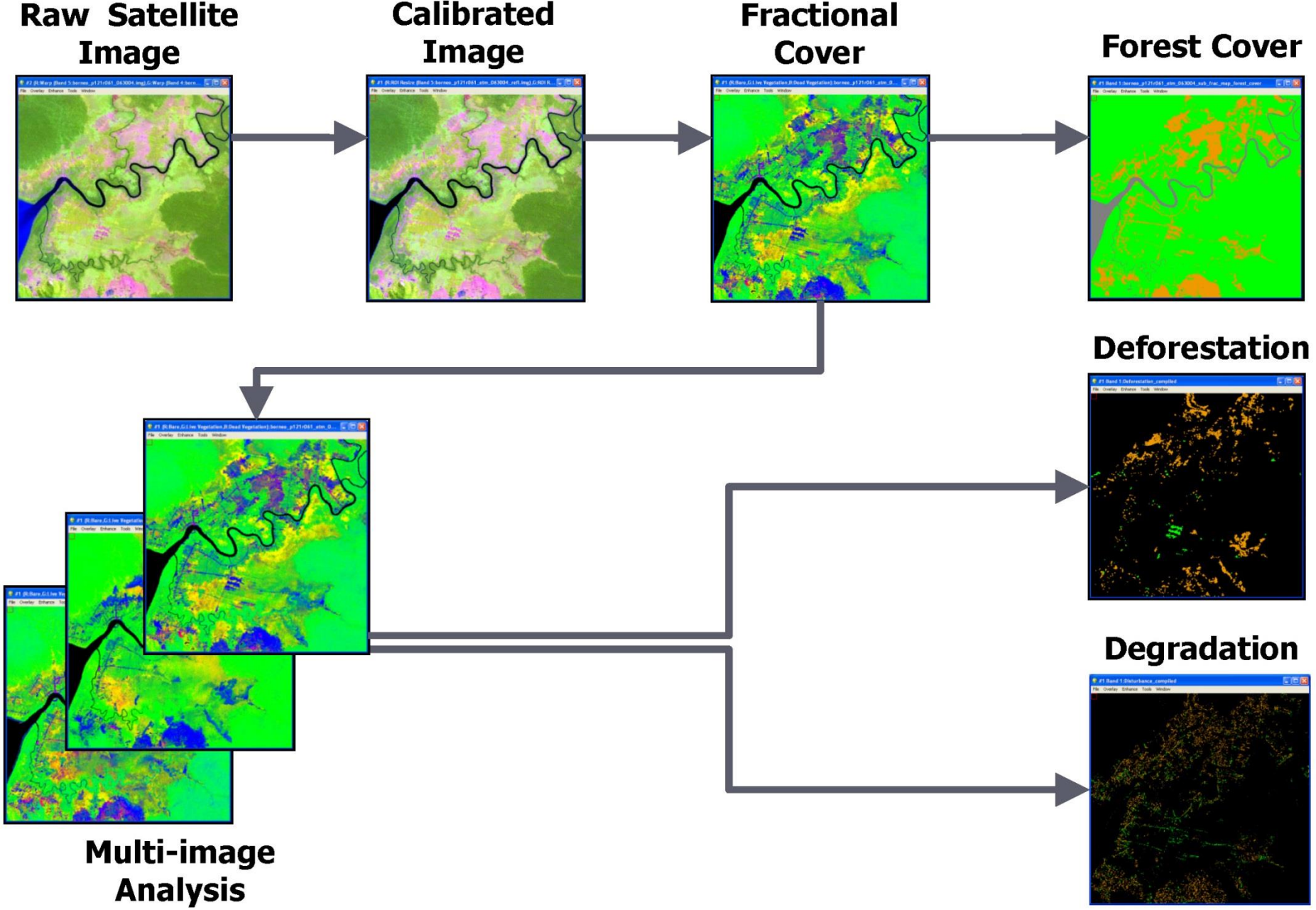
source: http://www.slideshare.net/Space-Applications/earth-observation-duevae


1972: Madrid as seen by Landsat-1:
60 m resolution; 4 spectral bands


2014: Madrid airport as seen by Worldview-3 (2014): 31cm resolution; 28 spectral bands

The “data explosion” in earth observation

- More satellites
- Higher resolution
- More data publicly available

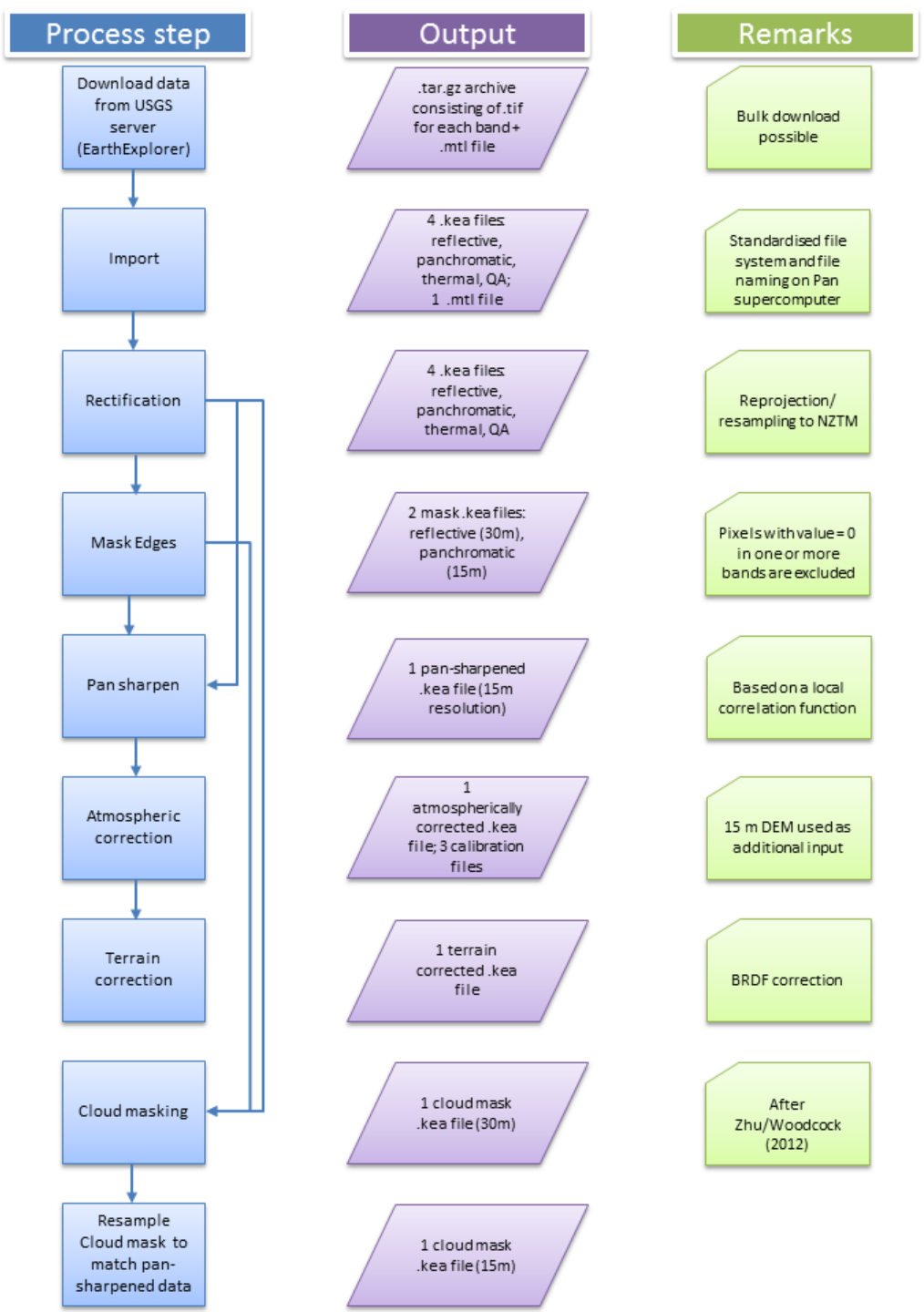


source: http://bunjilforestwatch.net/learn/forest-monitoring-sytems/remote-sensing/

Example 1: Automated processing of Landsat images

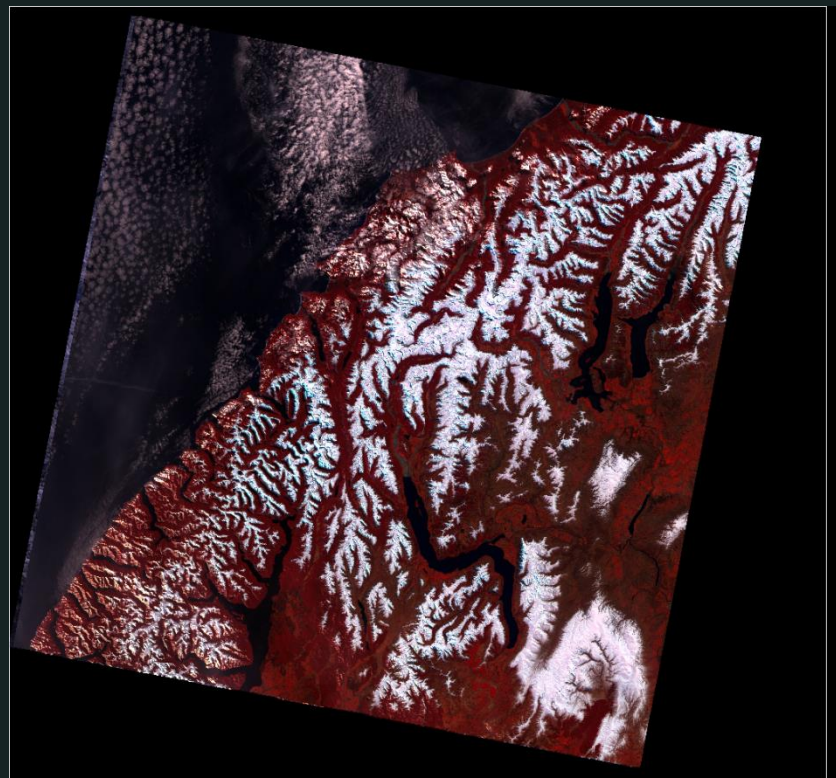
- Landsat series consists until now of 7 satellites (Landsat 1-5, 7-8; 6 didn't make it into orbit)
- Complete archive is made publicly available
- Longest complete record: from 1972 to now
- Mid-resolution (between 15 and 30 m) in the optical spectral range

Workflow


Process step
Output
Remarks
Download data from USGS server (EarthExplorer)
.tar.gz archive consisting of .tif for each band + .mtl file
Bulk download possible
Import
4 .kea files: reflective, panchromatic, thermal, QA; 1 .mtl file
Standardised file system and file naming on Pan supercomputer
Rectification
4 .kea files: reflective, panchromatic, thermal, QA
Reprojection/ resampling to NZTM
Mask Edges
2 mask .kea files: reflective (30m), panchromatic (15m)
Pixels with value = 0 in one or more bands are excluded
Pan sharpen
1 pan-sharpened .kea file (15m resolution)
Based on a local correlation function
Atmospheric correction
1 atmospherically corrected .kea file; 3 calibration files
15 m DEM used as additional input
Terrain correction
1 terrain corrected .kea file
BRDF correction
Cloud masking
1 cloud mask .kea file (30m)
After Zhu/Woodcock (2012)
Resample Cloud mask to match pan-sharpened data
1 cloud mask .kea file (15m)

Implementation details

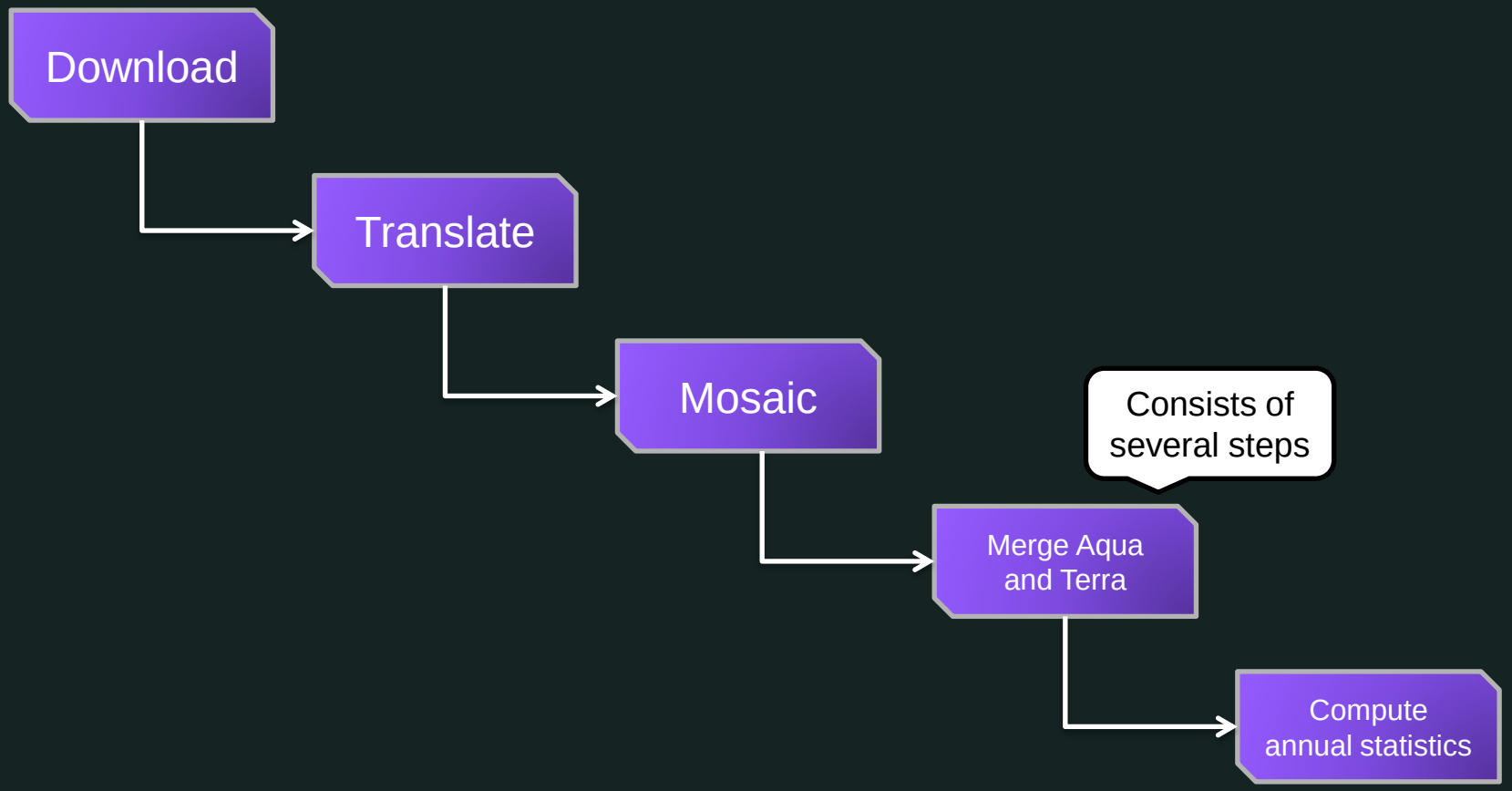
- 4252 landsat images processed so far; overall data volume 11.05 TB
- Batch level parallelization implemented using python, GDAL and SLURM

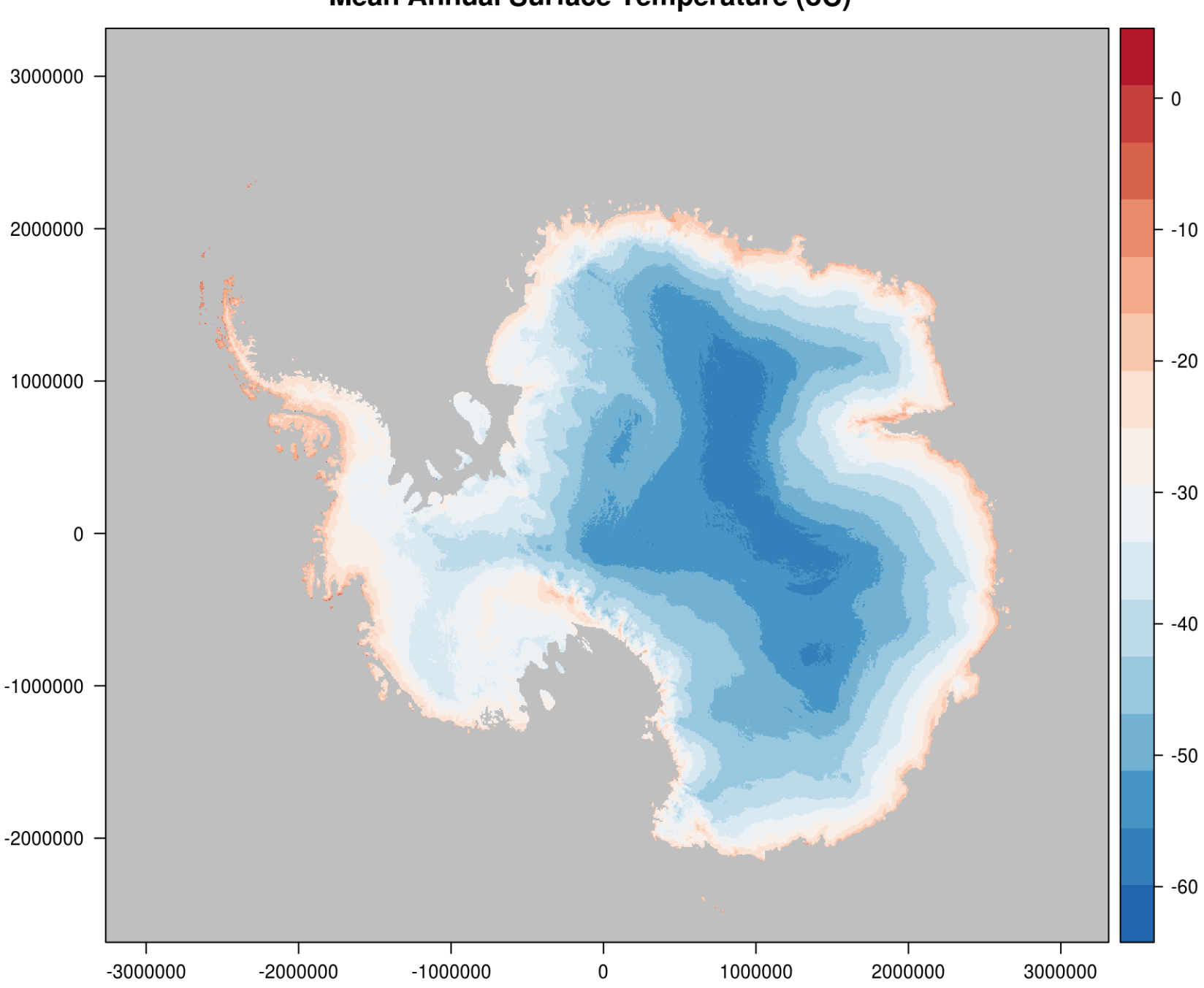


Example 2: MODIS Land Surface Temperature of the Antarctic

- MODIS programme consists of two satellites: Aqua and Terra
- Low spectral resolution: 1km
- Provides a number of derived products, e.g. land surface temperature (LST) datasets
- Complete time series from 2003 to 2014 was processed for the Antarctic
- Batch level parallelization using Python, GRASS and SLURM

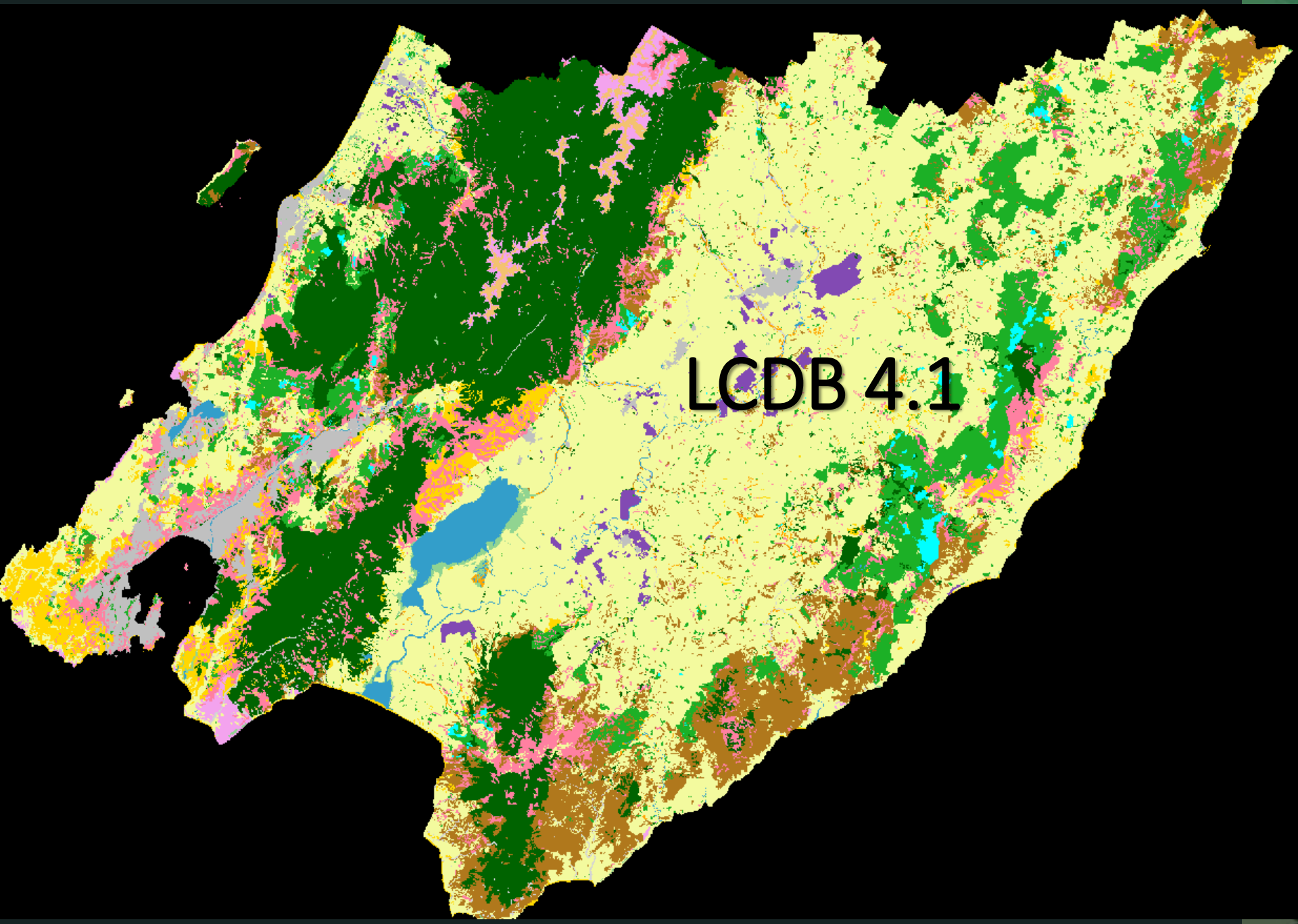
Workflow

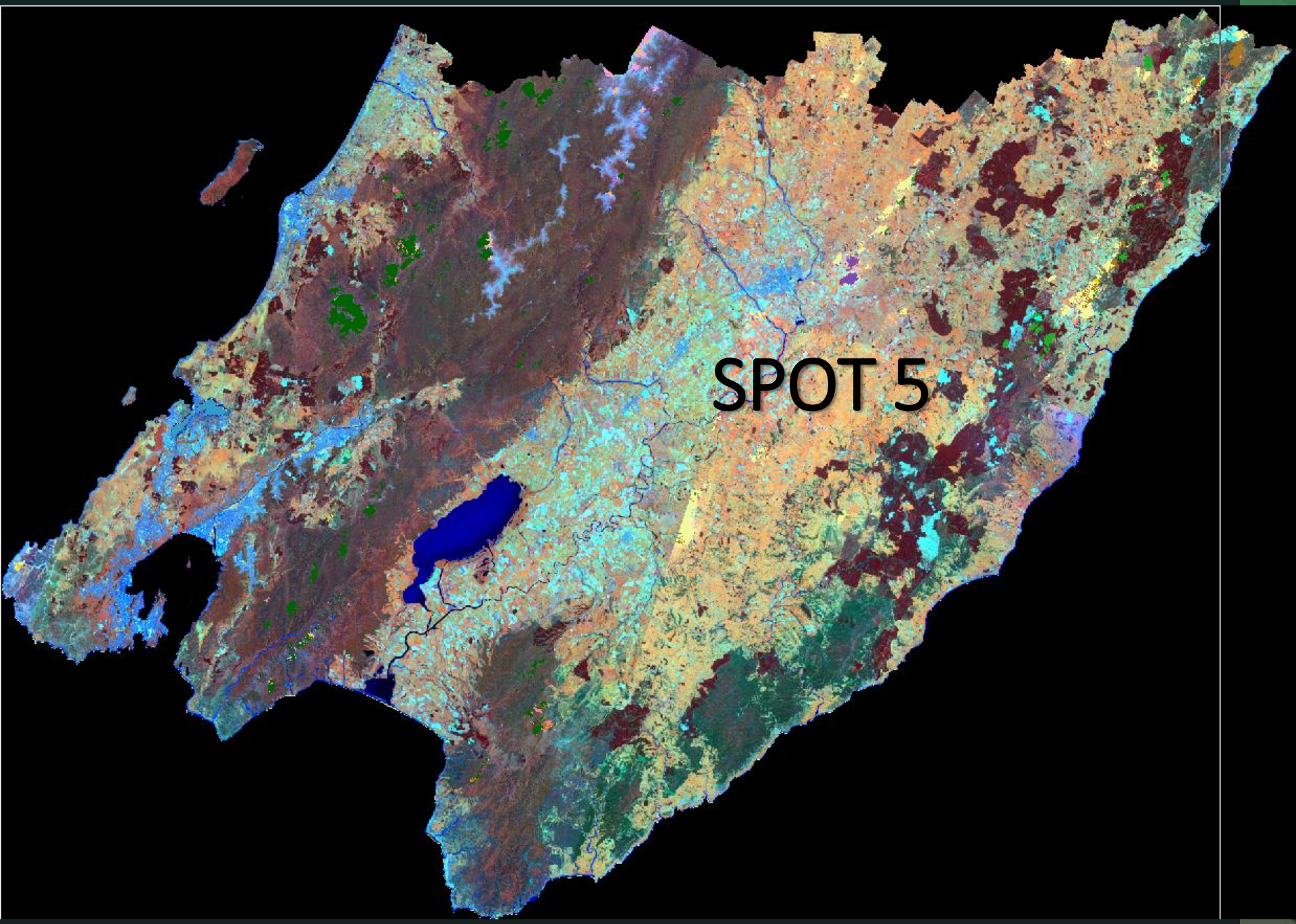

Download
Translate
Mosaic
Consists of several steps
Merge Aqua and Terra
Compute annual statistics


Mean Annual Surface Temperature (oC)
3000000
2000000
1000000
0
-1000000
-2000000
-3000000
-2000000
-1000000
0
1000000
2000000
3000000
0
-10
-20
-30
-40
-50
-60

Example 3: Improvement of the NZ Land cover database

- Land Cover DataBase (LCDB) of New Zealand is a multi-temporal, digital thematic map of land cover and land use
- Mainly created by manual interpretation and digitization of SPOT satellite images
- Many small woody patches on pasture background are unmapped – this is what we wanted to improve

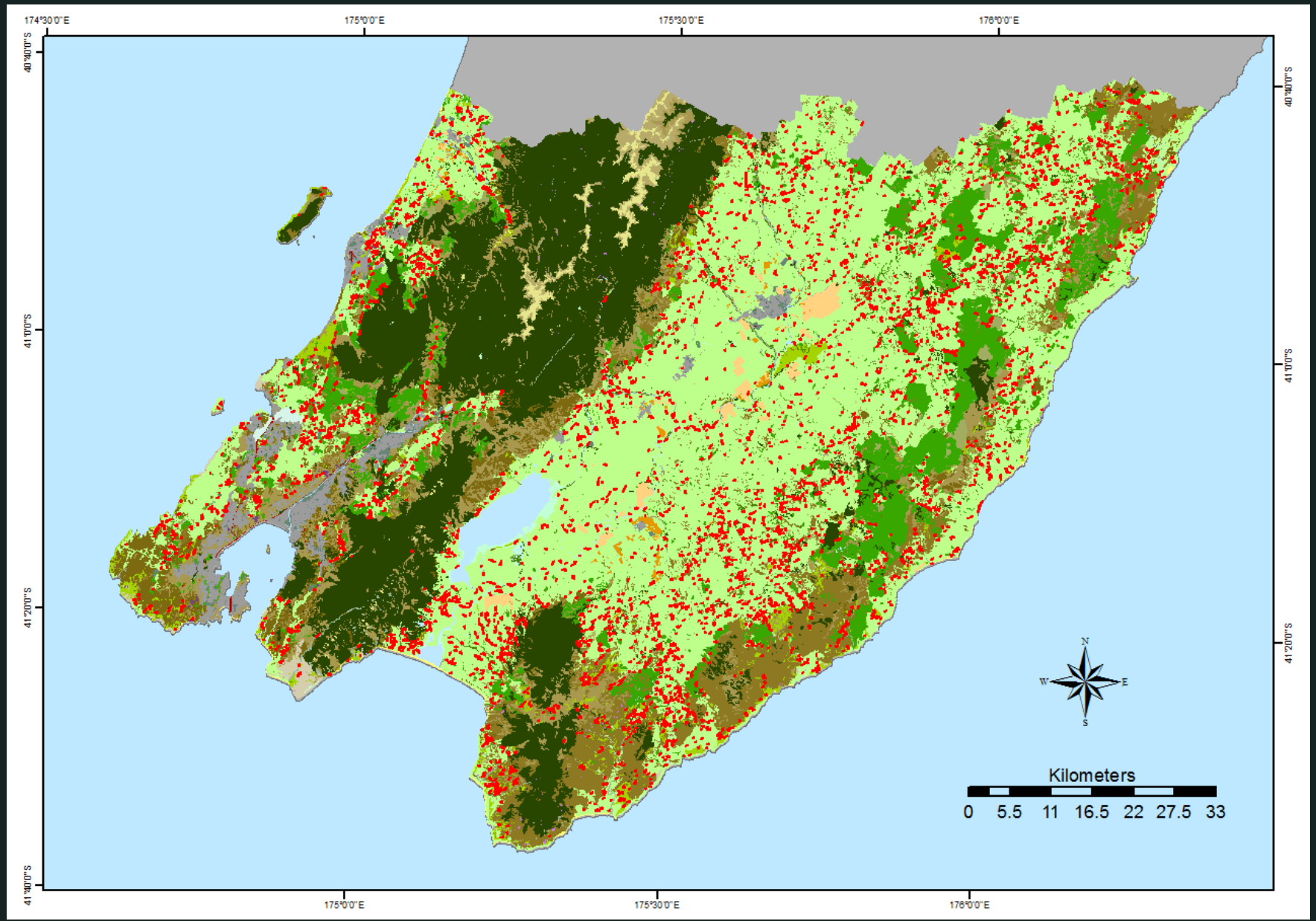

LCDB 4.1

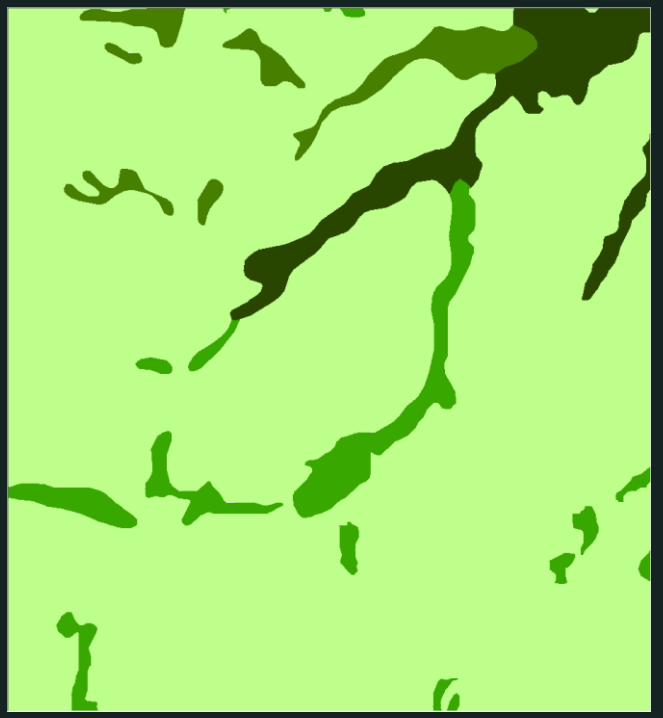

SPOT 5

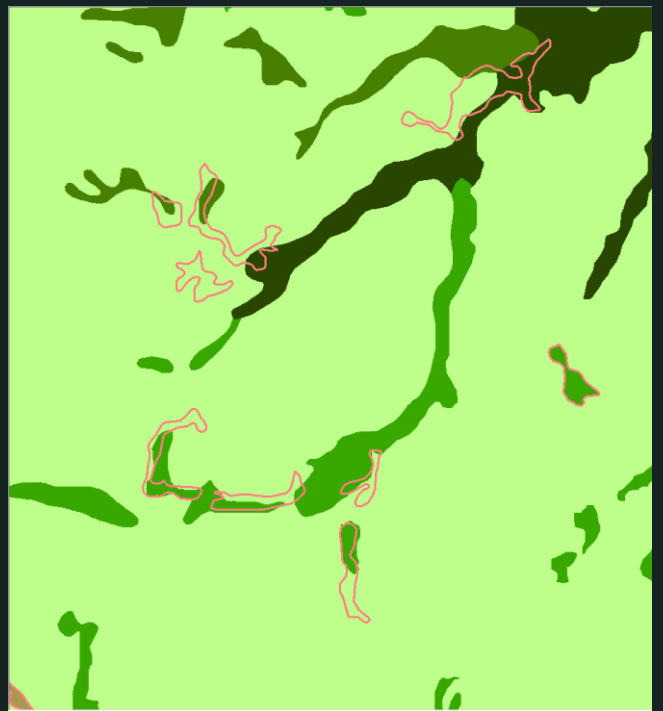

PALSAR


LiDAR CHM

Woody patches Wellington region







Results

Region	Woody Sites	Total Area (ha)
Northland	10271/1820	**21478/2892**
Auckland	4430/1124	**8865/1771**
Waikato	5753/1565	**10635/2441**
Bay of Plenty	2443/466	**5043/749**
Gisborne	4168/1120	**8727/1950**
Hawke's Bay	5235/1198	**11060/1978**
Manawatu-Wanganui	7437/2366	**14335/3831**
Taranaki	2775/741	**5492/1197**
Wellington	4139/1473	**8033/2408**
Tasman	2237/230	**4889/399**
Nelson	105/14	**201/22**
Marlborough	1647/394	**3568/688**
Canterbury	6461/2153	**13482/3644**
West Coast	1619/155	**3704/262**
Otago	2928/1011	**5744/1625**
Southland	2629/1010	**5766/1688**
Total	**64277/16840**	**131022/27545**

Implementation details

- Batch level parallelization (based on NZ regions) implemented using python, GDAL and SLURM
- Support Vector Classification using scikit-learn

Conclusion

- Processing and analysis of satellite images is a “Big Data” problem
- The NeSI Pan cluster helps us immensely in our analyses, but…
- requirements are different from other use cases:
  - Operations are very I/O intensive
  - For most problems batch level parallelism offers the best solution
  - Need to store large amounts of data

Thank you!