Satellite time-series processing with FORCE in CSC *(Taito)* environment

FORCE

Framework for Operational Radiometric Correction for Environmental Monitoring

Pekka Hurskainen (pekka.hurskainen@ymparisto.fi)

Finnish Environment Institute SYKE

University of Helsinki, Dept. of Geosciences and Geography, ECHOLAB

SYKE

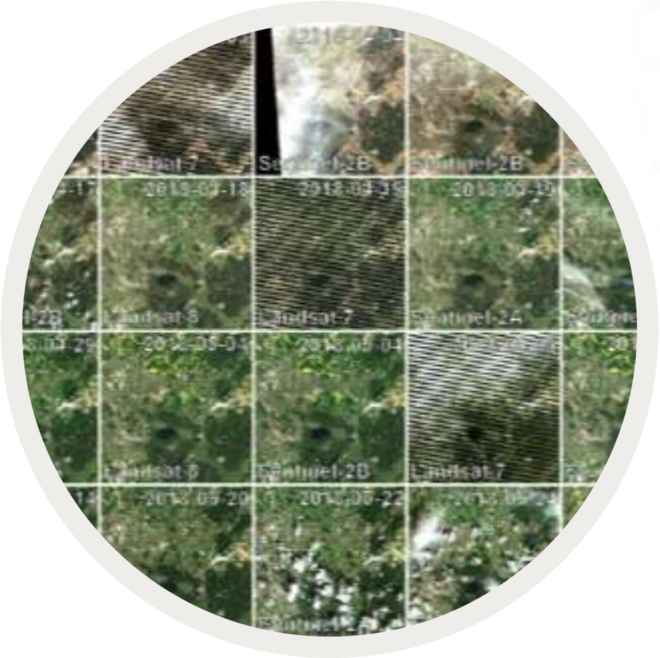
FORCE

David Frantz

Humboldt-Universität zu Berlin | HU Berlin · Department of Geography

25.96 · Dr. rer. nat.

FORCE is a software project, which is intended to be a Landsat+Sentinel-2 'all-in-one solution' for generating atmospherically and topographically corrected, quality screened images, image composites and basic time series products tailored for large-area and multi-temporal applications.



Force homepage:

https://www.uni-trier.de/index.php?id=63673&L=0

SYKE

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

An Operational Radiometric Landsat Preprocessing Framework for Large-Area Time Series Applications

David Frantz, Achim Röder, Marion Stellmes, and Joachim Hill

ELSEVIER

Remote Sensing of Environment

Volume 190, 1 March 2017, Pages 331-347

Phenology-adaptive pixel-based compositing using optical earth observation imagery

David Frantz [a], Achim Röder [a], Marion Stellmes [a, b], Joachim Hill [a]

Show more

https://doi.org/10.1016/j.rse.2017.01.002

remote sensing

MDPI

Technical Note

FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond

David Frantz

Earth Observation Lab, Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany; david.frantz@geo.hu-berlin.de

Received: 11 April 2019; Accepted: 28 April 2019; Published: 10 May 2019

check for updates

SYKE

Main features 1/3

- Written in C, parallelization with OpenMP API. Command-line!
- **Integration** of Landsat 4–8, and Sentinel-2 A/B as Virtual Constellation
- **Data management** of L4-8 and S2 Level 1 data, download
- Generation of **Analysis Ready Data** (ARD) a.ka Level2 data
  - Advanced cloud and cloud shadow detection (`modified` **`fmask`**)
  - Quality screening
  - Integrated atmospheric[1] and topographic[2] correction
  - Adjacency effect correction
  - BRDF reduction
  - Resolution merge of Sentinel-2 bands: 20m -> 10m
  - Data cubing: reprojection / gridding

[1]Radiative transfer model of Tanré et al 1990
[2]Modified c-correction of Kobayashi & Sanga-Ngoie 2008

SYKE

Main features 2/3

- Generation of **highly Analysis Ready Data** (hARD) a.ka Level3
  - Large area, Gap free
  - Best Available Pixel (BAP) composites
  - Phenology Adaptive Composites (PAC)
  - Spectral Temporal Metrics (STM) – min, max, mean, range, quantiles…
- Generation of **highly Analysis Ready Data plus** (hARD+) a.ka Level4
  - Time Series generation: spectral bands, spectral indices, Spectral Mixture Analysis (SMA)
  - Time series folding
  - Time series interpolation
  - Time series statistics
  - Trend analysis
  - Change, Aftereffect, Trend (CAT) analysis
  - Land Surface Phenology (LSP)

SYKE

Main features 3/3

- Detailed **data mining** of the **Clear Sky Observation** (CSO) availability
- **Data Fusion**. Improving spatial resolution of coarse continuous fields:
  - MODIS LSP -> medium resolution LSP.
  - Improving spatial resolution of lower resolution ARD using higher resolution ARD: 30m Landsat -> 10m using Sentinel-2 targets

SYKE

Data Cubes in FORCE

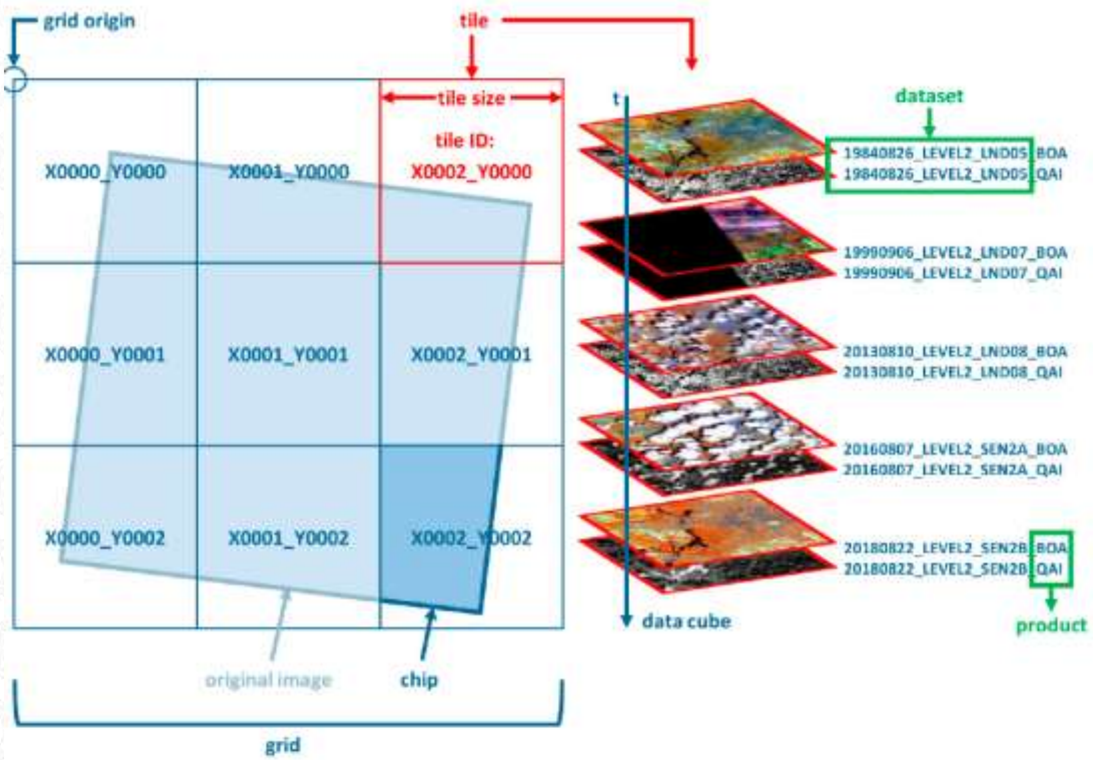
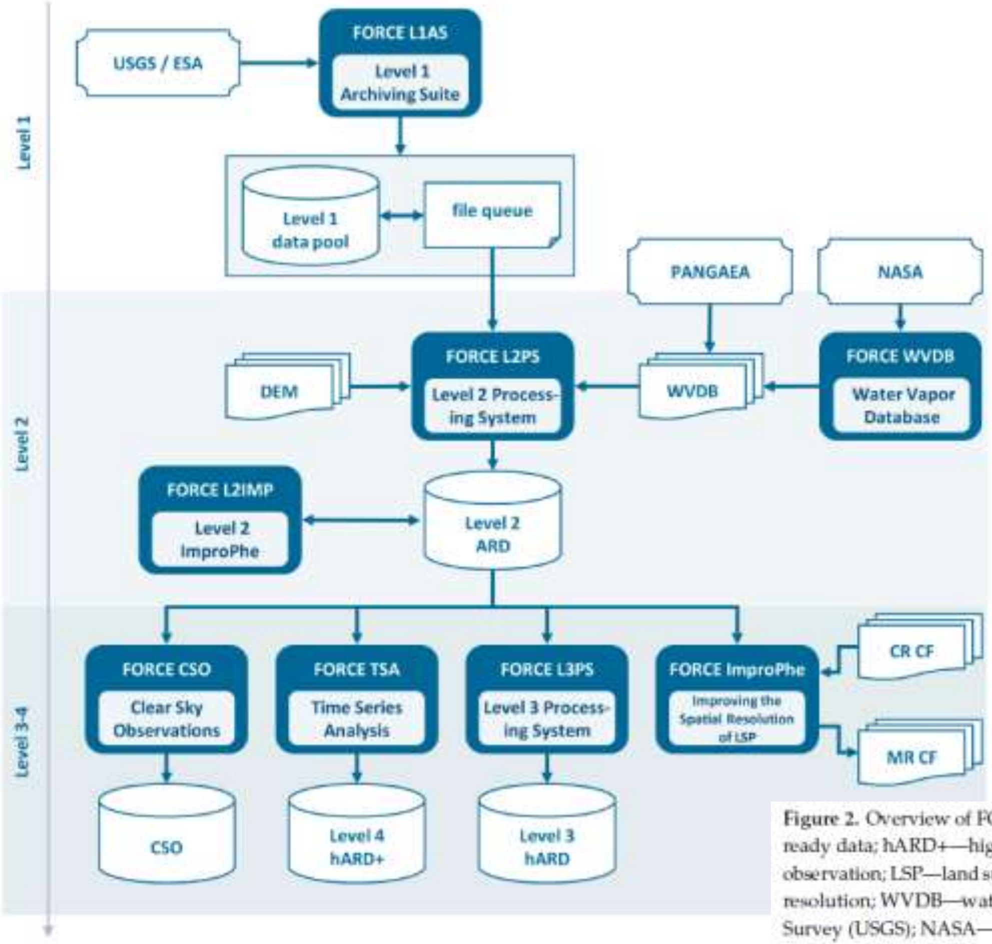


Figure 1. Overview of gridding and data cube terminology used in FORCE (Framework for Operational Radiometric Correction for Environmental monitoring).

Frantz 2019

SYKE

General workflow



Frantz 2019

Figure 2. Overview of FORCE, general workflow. ARD—analysis ready data; hARD—highly analysis ready data; hARD+—highly analysis ready data plus; DEM—digital elevation model; CSO—clear sky observation; LSP—land surface phenology; CF—continuous field; CR—coarse resolution; MR—medium resolution; WVDB—water vapor database; ESA—European Space Agency; USGS—U.S. Geological Survey (USGS); NASA—National Aeronautics and Space Administration.

S Y K E

FORCE in Taito

- Installed to Taito in December 2018 (v2.1)
- Load modules:

`module load force gnu-parallel`

- List available commands:

`ls -l /appl/earth/force/v2.1/bin`

- Note, some commands have `.sh` extension some don't!

```
[pjhurska@c308 kili]$ ls -l /appl/earth/force/v2.1/bin
total 4828
-rwxrwxr-x 1 ejarvine csc   8568 Dec  4 16:22 force
-rwxrwxr-x 1 ejarvine csc 475864 Dec  4 16:22 force-cso
-rwxrwxr-x 1 ejarvine csc 477344 Dec  4 16:22 force-improphe
-rwxrwxr-x 1 ejarvine csc 479128 Dec  4 16:22 force-l2imp
-rwxrwxr-x 1 ejarvine csc 474992 Dec  4 16:22 force-l2ps
-rwxrwxr-x 1 ejarvine csc  11018 Dec  4 16:22 force-level1-landsat.sh
-rwxrwxr-x 1 ejarvine csc   6767 Dec 10 10:49 force-level1-sentinel2-long.sh
-rwxrwxr-x 1 ekkylli  csc   8211 Dec 12 12:53 force-level1-sentinel2_nv.sh
-rwxrwxr-x 1 ejarvine csc   8209 Dec 10 10:49 force-level1-sentinel2.sh
-rwxrwxr-x 1 ejarvine csc   4717 Dec 19 09:03 force-level2.sh
-rwxrwxr-x 1 ejarvine csc 481408 Dec  4 16:22 force-level3
-rwxrwxr-x 1 ejarvine csc 474120 Dec  4 16:22 force-lut-modis
-rwxrwxr-x 1 ejarvine csc   2450 Dec  4 16:22 force-mosaic.sh
-rwxrwxr-x 1 ejarvine csc  11480 Dec  4 16:22 force-parameter-cso
-rwxrwxr-x 1 ejarvine csc  11256 Dec  4 16:22 force-parameter-improphe
-rwxrwxr-x 1 ejarvine csc  11096 Dec  4 16:22 force-parameter-l2imp
-rwxrwxr-x 1 ejarvine csc  12000 Dec  4 16:22 force-parameter-level2
-rwxrwxr-x 1 ejarvine csc  14584 Dec  4 16:22 force-parameter-level3
-rwxrwxr-x 1 ejarvine csc  13992 Dec  4 16:22 force-parameter-tsa
-rwxrwxr-x 1 ejarvine csc 476096 Dec  4 16:22 force-qai-inflate
-rwxrwxr-x 1 ejarvine csc   1970 Dec  4 16:22 force-quicklook-level2.sh
-rwxrwxr-x 1 ejarvine csc   2472 Dec  4 16:22 force-quicklook-level3.sh
-rwxrwxr-x 1 ejarvine csc 477872 Dec  4 16:22 force-tabulate-grid
-rwxrwxr-x 1 ejarvine csc 472936 Dec  4 16:22 force-tile-finder
-rwxrwxr-x 1 ejarvine csc 478408 Dec  4 16:22 force-tsa
[pjhurska@c308 kili]$
```

SYKE

FORCE L1AS - Level 1 Archiving Suite

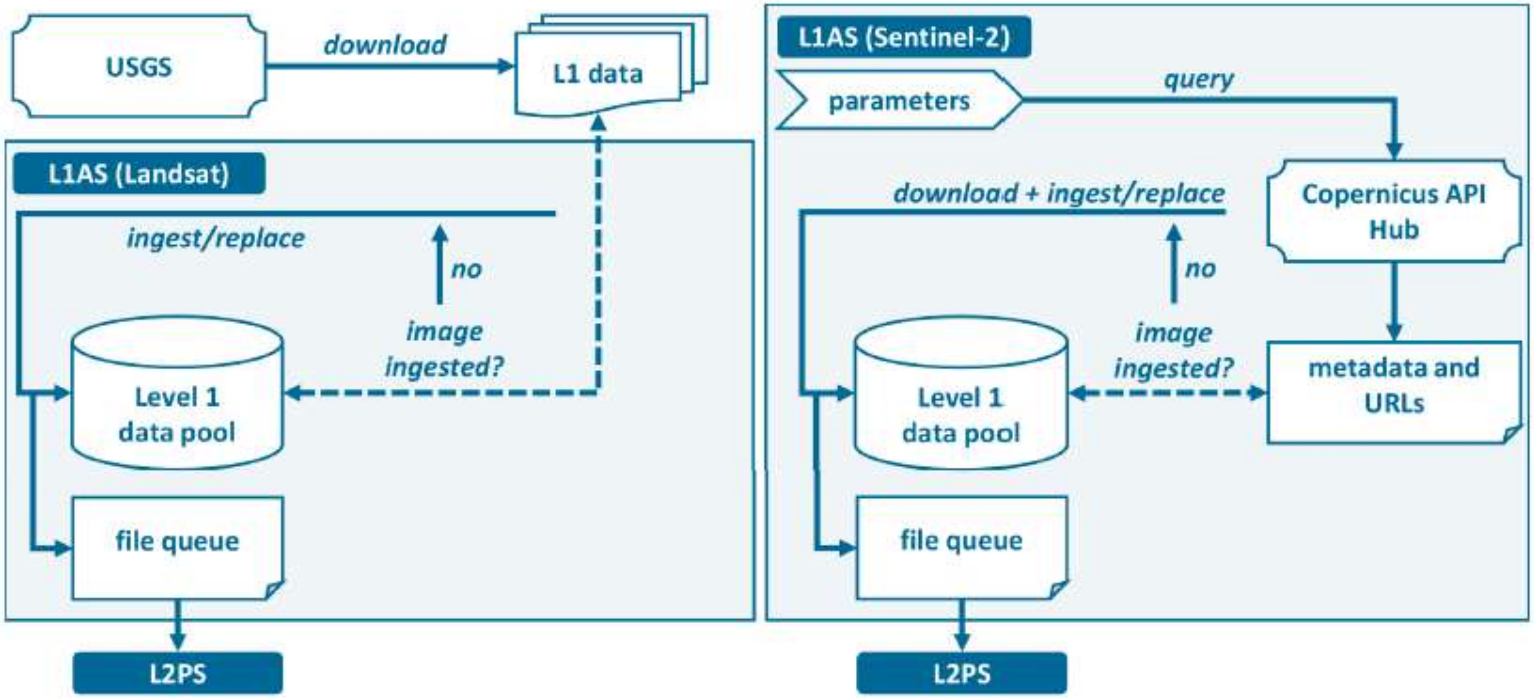


Figure 3. FORCE Level 1 Archiving Suite (L1AS) workflow.

Frantz 2019

Example: Download 3+ years of S2 images from Kilimanjaro, Tanzania

Module	force-level1-sentinel2 + force-level1-sentinel2-long
Usage	force-level1-sentinel2 Level-1-datapool queue boundingbox starttime endtime min-cc max-cc

- `force-level1-sentinel2-nv.sh /wrk/pjhurska/kili level1.pool "37.841/-3.58,36.97/-3.58,36.97/-2.805,37.841/-2.805,37.841/-3.58" 2016-01-01 2019-02-01 0 85`
- Downloads all available sentinel-2 images from the area of interest, from 1 Jan 2016 until 1 Feb 2019, allowed cloud coverage 0-85%
- You need user account at scihub.copernicus.eu (user name and password in a hidden file '`.scihub`' in your home directory)

SYKE

FORCE L2PS - Level 2 Processing System

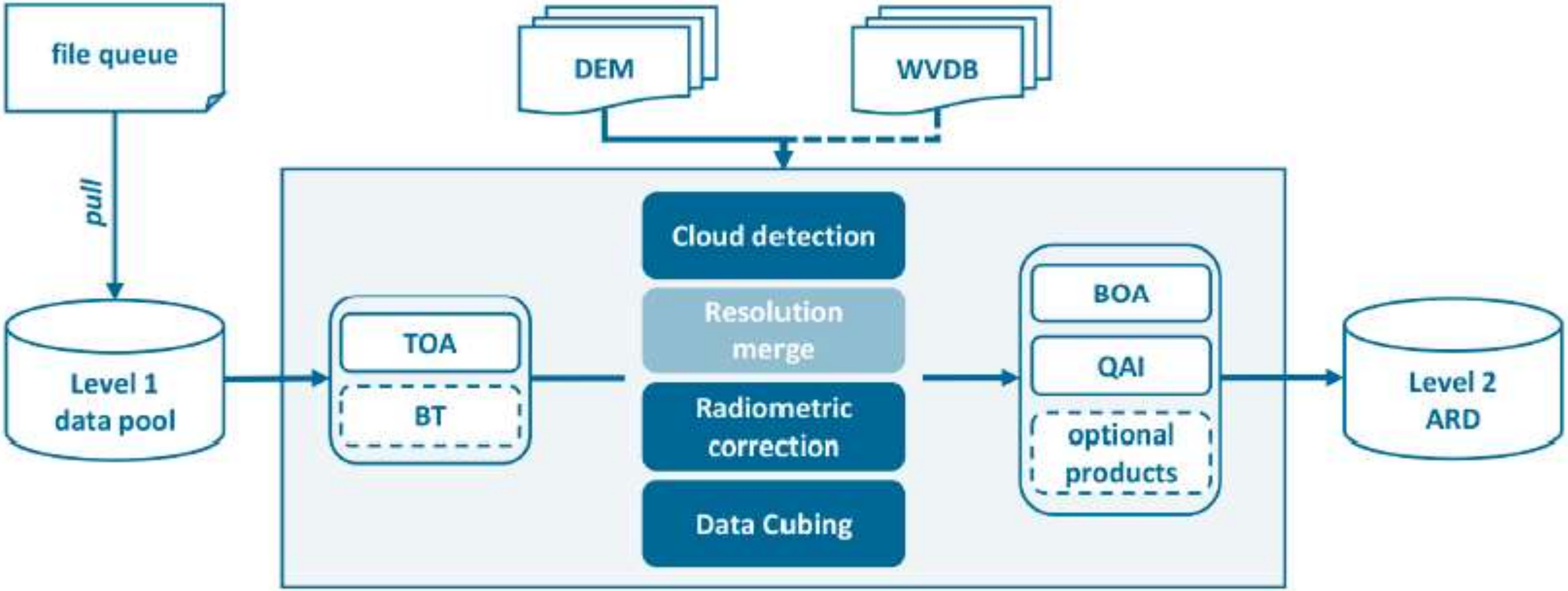


Figure 4. FORCE Level 2 Processing System (L2PS) workflow. TOA—top-of-atmosphere; BT—brightness temperature; BOA—bottom-of-atmosphere; QAI—quality assurance information.

Frantz 2019

SYKE

force-level2 – mass processing of level1 images

Module	force-level2
Usage	force-level2 par-file ncpu delay

- After parameters are set, run the actual level2 processing **as a batch job**
- Arguments: (1) name of the parameter file, (2) number of CPUs used for parallel processing, (3) delay in seconds between image processing
- After processing, the image is dequeued -> if you need to reprocess your images, you can `sed -i 's/DONE/QUEUED/' level1.pool`
- The processed images and a logfile are written to output dir

SYKE

Example of a SLURM batch file – run-level2.sh

```
#!/bin/bash -l
# created: Dec 16, 2018 11:13 PM
# author: pjhurska
#SBATCH -J run-level2
#SBATCH --constraint="snb|hsw"
#SBATCH -o level2.out
#SBATCH -e level2.err
#SBATCH -p serial
#SBATCH -N 1
#SBATCH -n 1
#SBATCH --cpus-per-task=10
#SBATCH -t 08:00:00
#SBATCH --mem=120000
#SBATCH --mail-type=END
#SBATCH --mail-user=pekka.hurskainen@helsinki.fi

srun force-level2.sh level2.prm 10 20

# This script will print some usage statistics to the
# end of file: level2.out
# Use that to improve your resource request estimate
# on later jobs.
used_slurm_resources.bash

```

S Y K E

Run the batch file

- Schedule the batch job for processing:
  - `sbatch run-level2.sh`
- Check the status of your batch queue
  - `squeue -u pjhurska`
- You will be notified by email after the job is done, or crashed in an error (settings for that in the SLURM batch file)

SYKE

Example of a SLURM "standard output"

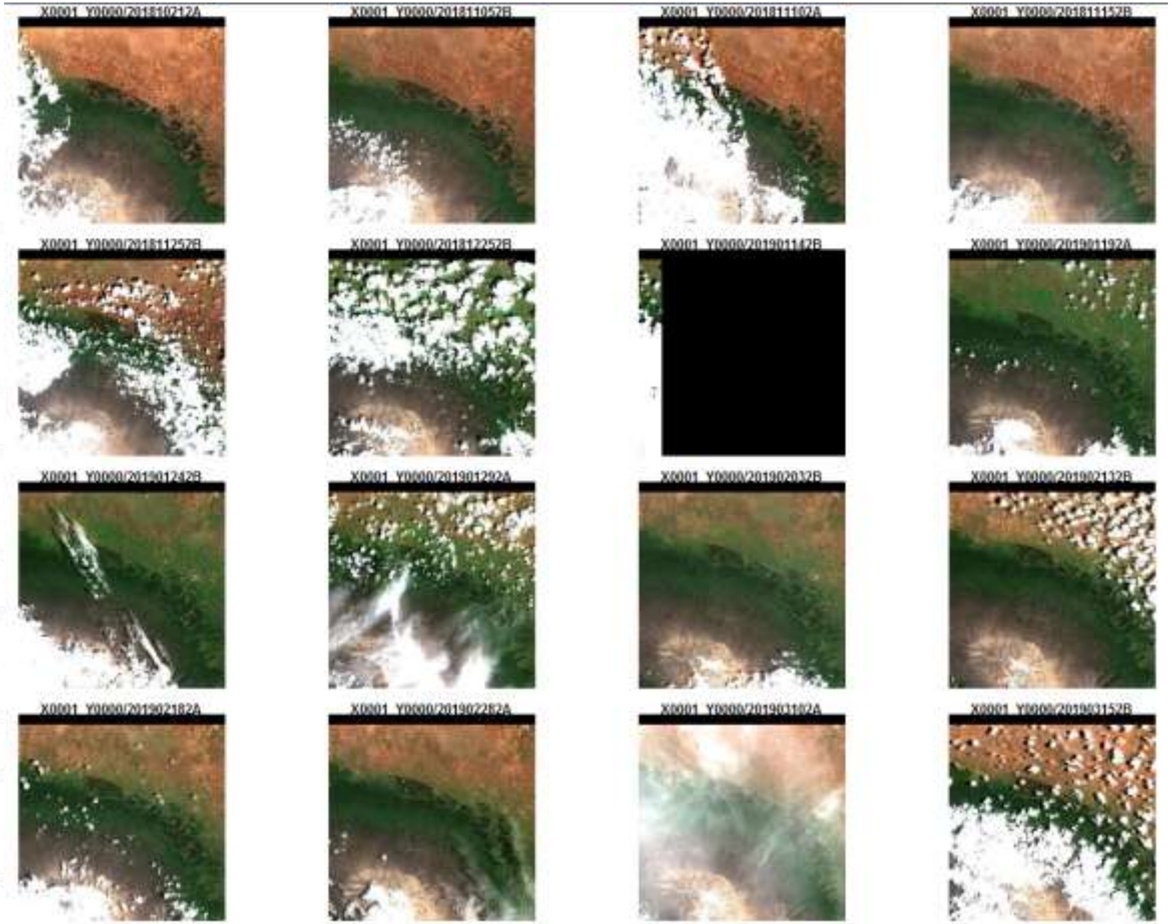
```
87 images enqueued. Start processing with 10 CPUs
Job ID: 37779858
Cluster: csc
User/Group: pjhurska/pjhurska
State: RUNNING
Nodes: 1
Cores per node: 10
CPU Utilized: 19:10:10
CPU Efficiency: 81.77% of 23:26:40 core-walltime
Memory Utilized: 109.89 GB
Memory Efficiency: 93.77% of 117.19 GB
Job consumed 140.67 CSC billing units based on memory reservation multiplier
WARNING: Efficiency statistics may be misleading for RUNNING jobs.
```

Structure of the Level2 Data Cube

- Each datacube has the following components created automatically:
  - Subfolders for each tile (e.g. X0000_Y0000, X0000_Y0001, … Xnnnn_Ynnnn). Under each subfolder, output images for each tile and for each image that was processed.
  - Available output products, in GeoTIFF or ENVI format:

BOA	Bottom-of-Atmosphere Reflectance (standard output, scale: 10000, nodata: -9999)
TOA	Top-of-Atmosphere Reflectance (secondary standard output, scale: 10000, nodata: -9999)
QAI	Quality Assurance Information (standard output, bit coding, nodata: 1)
AOD	Aerosol Optical Depth (550 nm, optional output, scale: 1000, nodata: -9999)
CLD	Cloud / Cloud shadow distance (optional output, scale: 10000, nodata: -9999)
WVP	Water vapor (optional output, scale: 1000, nodata: -9999)
VZN	View zenith (optional output, scale: 100, nodata: -9999)
HOT	Haze Optimized Transformation (optional output, scale: 10000, nodata: -9999)

Quicklooks of one tile in the data cube



SYKE

FORCE L3PS - Level 3 Processing System

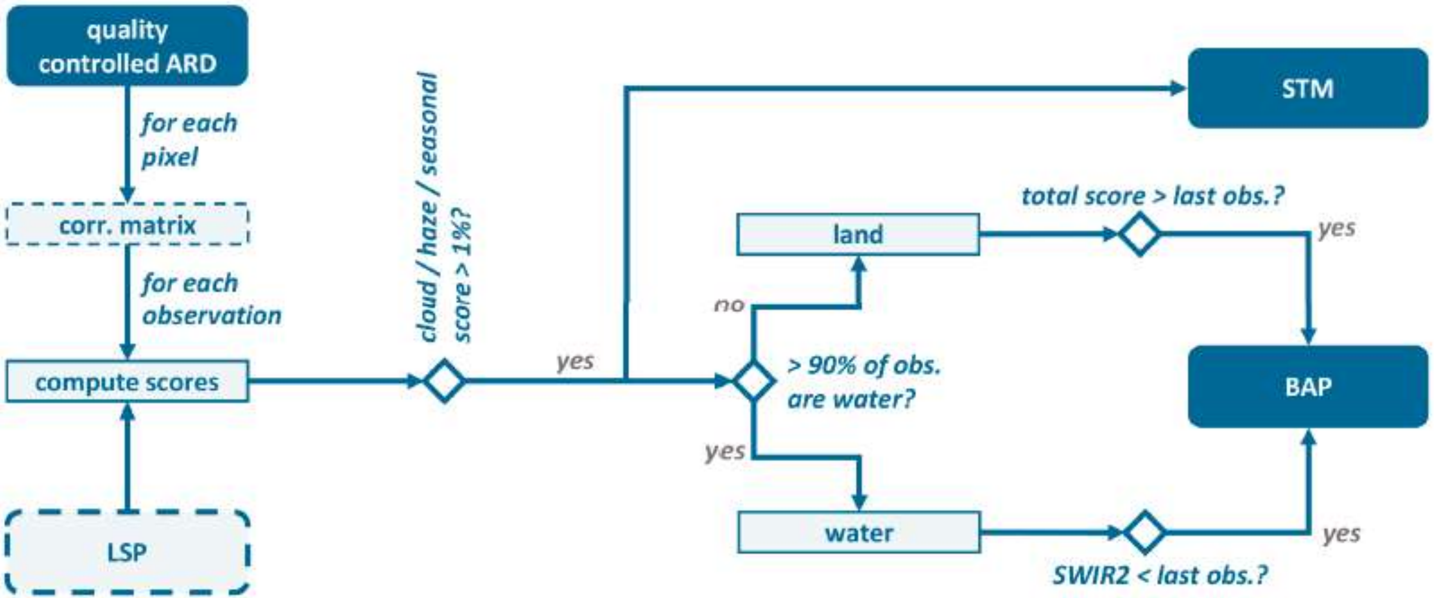


Figure 8. Processing workflow of the FORCE Level 3 Processing System (L3PS). STM—spectral temporal metric; BAP—best-available pixel composite.

SYKE

Example of a SLURM batch file – run-level3.sh

```
#!/bin/bash -l
# created: Dec 16, 2018 11:13 PM
# author: pjhurska
#SBATCH -J run-level3
#SBATCH --constraint="snb|hsw"
#SBATCH -o level3.out
#SBATCH -e level3.err
#SBATCH -p serial
#SBATCH -N 1
#SBATCH -n 1
#SBATCH --cpus-per-task=16
#SBATCH -t 04:00:00
#SBATCH --mem=100000
#SBATCH --mail-type=END
#SBATCH --mail-user=pekka.hurskainen@helsinki.fi

srun force-level3 level3.prm

# This script will print some usage statistics to the
# end of file: s2test_out
# Use that to improve your resource request estimate
# on later jobs.
used_slurm_resources.bash
```

- Schedule the batch job for processing:
  - `sbatch run-level3.sh`

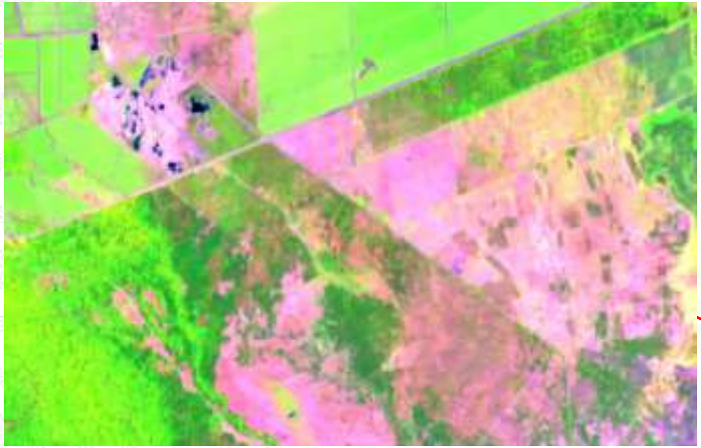
SYKE

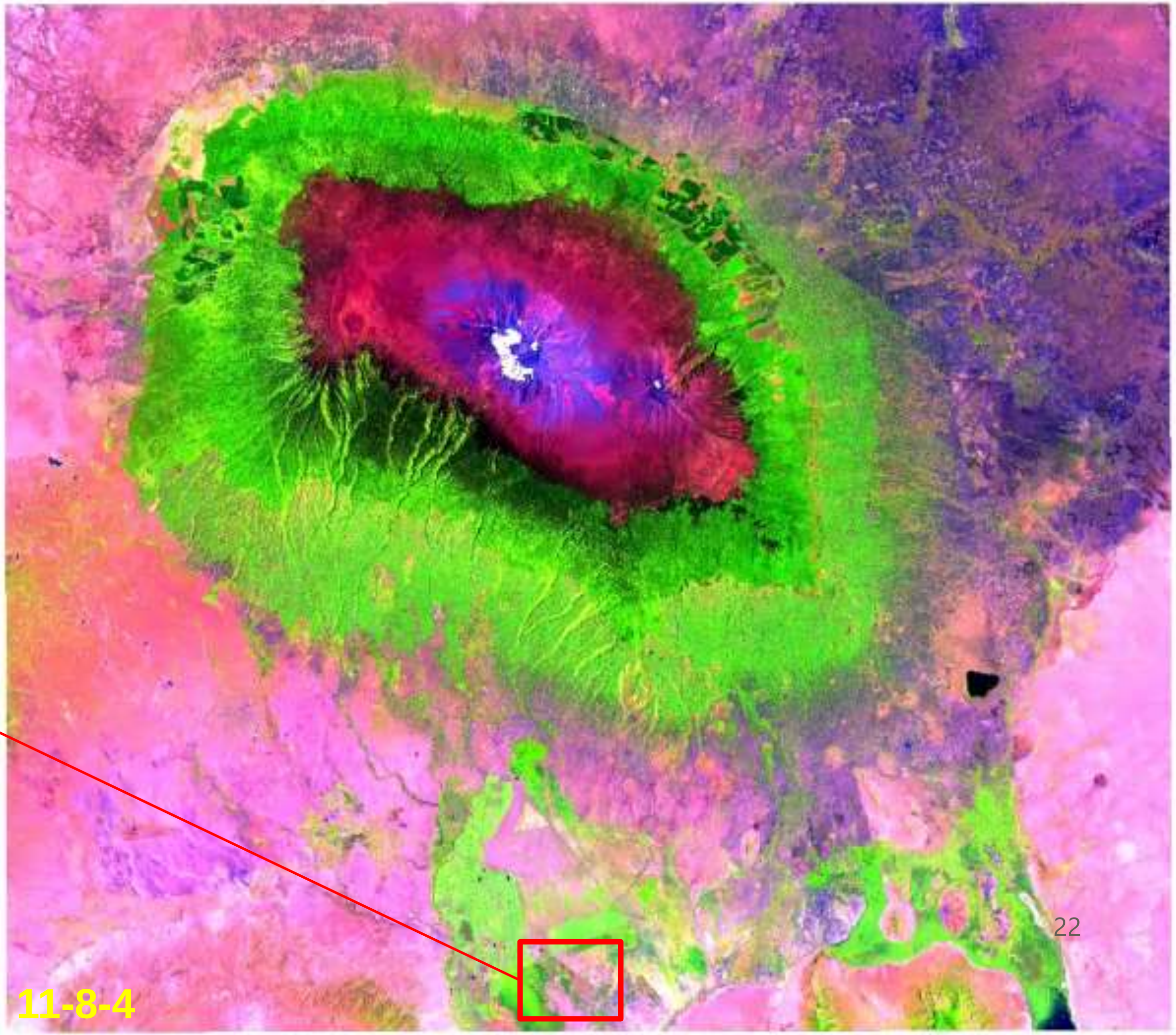
Level 3 outputs

BAP	Best Available Pixel composite (optional output, scale: 10000, nodata: -9999)
INF	Compositing Information (optional output, nodata: 1/-9999)
SCR	Compositing Scores (optional output, scale: 10000, nodata: -9999)
AVG	Temporal Average (optional output, scale: 10000, nodata: -9999)
STD	Temporal Standard Deviation (optional output, scale: 10000, nodata: -9999)
MIN	Temporal Minimum (optional output, scale: 10000, nodata: -9999)
MAX	Temporal Maximum (optional output, scale: 10000, nodata: -9999)
RNG	Temporal Range (optional output, scale: 10000, nodata: -9999)
SKW	Temporal Skewness (optional output, scale: 10000, nodata: -9999)
KRT	Temporal Kurtosis (optional output, scale: 10, nodata: -9999)
Q25	Temporal 0.25 Quantile (optional output, scale: 10000, nodata: -9999)
Q50	Temporal 0.50 Quantile (optional output, scale: 10000, nodata: -9999)
Q75	Temporal 0.75 Quantile (optional output, scale: 10000, nodata: -9999)
IQR	Temporal Interquartile Range (optional output, scale: 10000, nodata: -9999)

SYKE

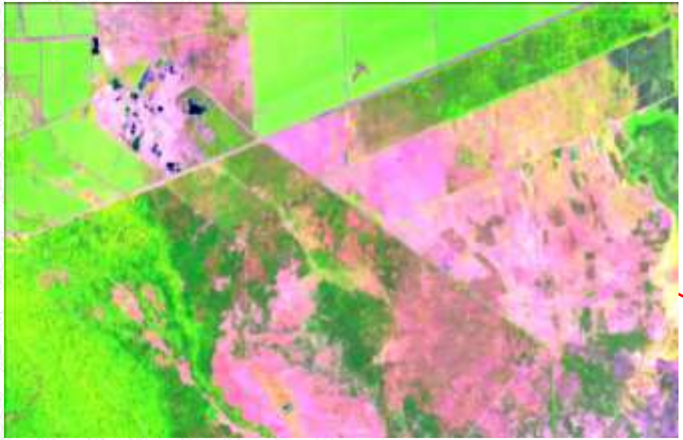
- Temporal average surface reflectance 01/2016 – 09/ 2019
- All composites in 10m pixel size

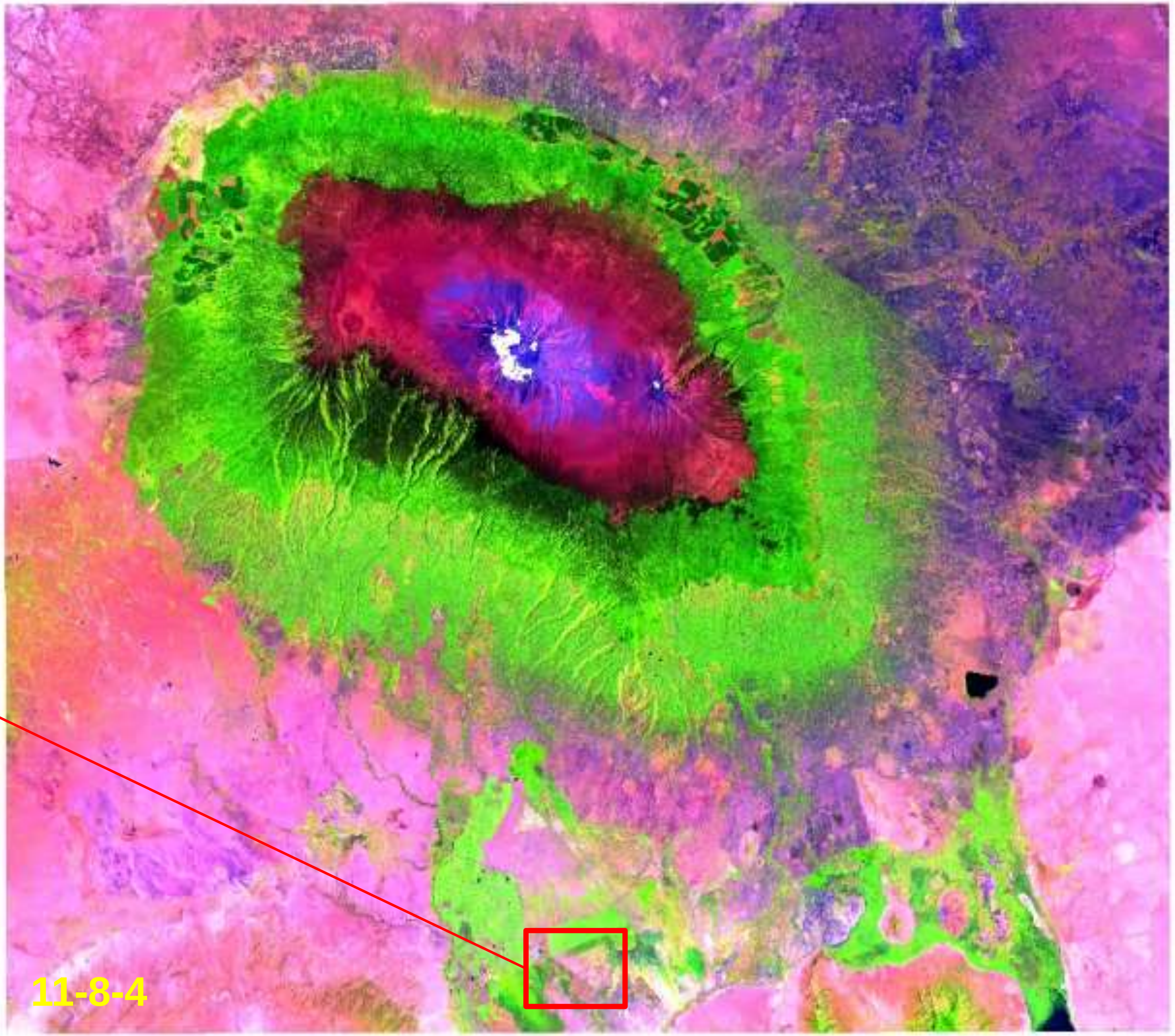




SYKE

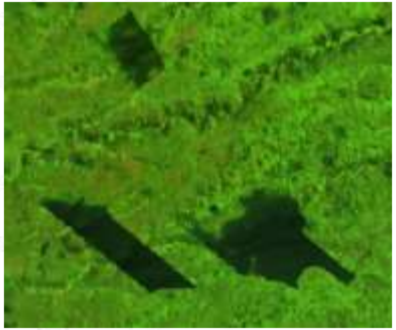
- Temporal median surface reflectance 01/2016 – 09/ 2019





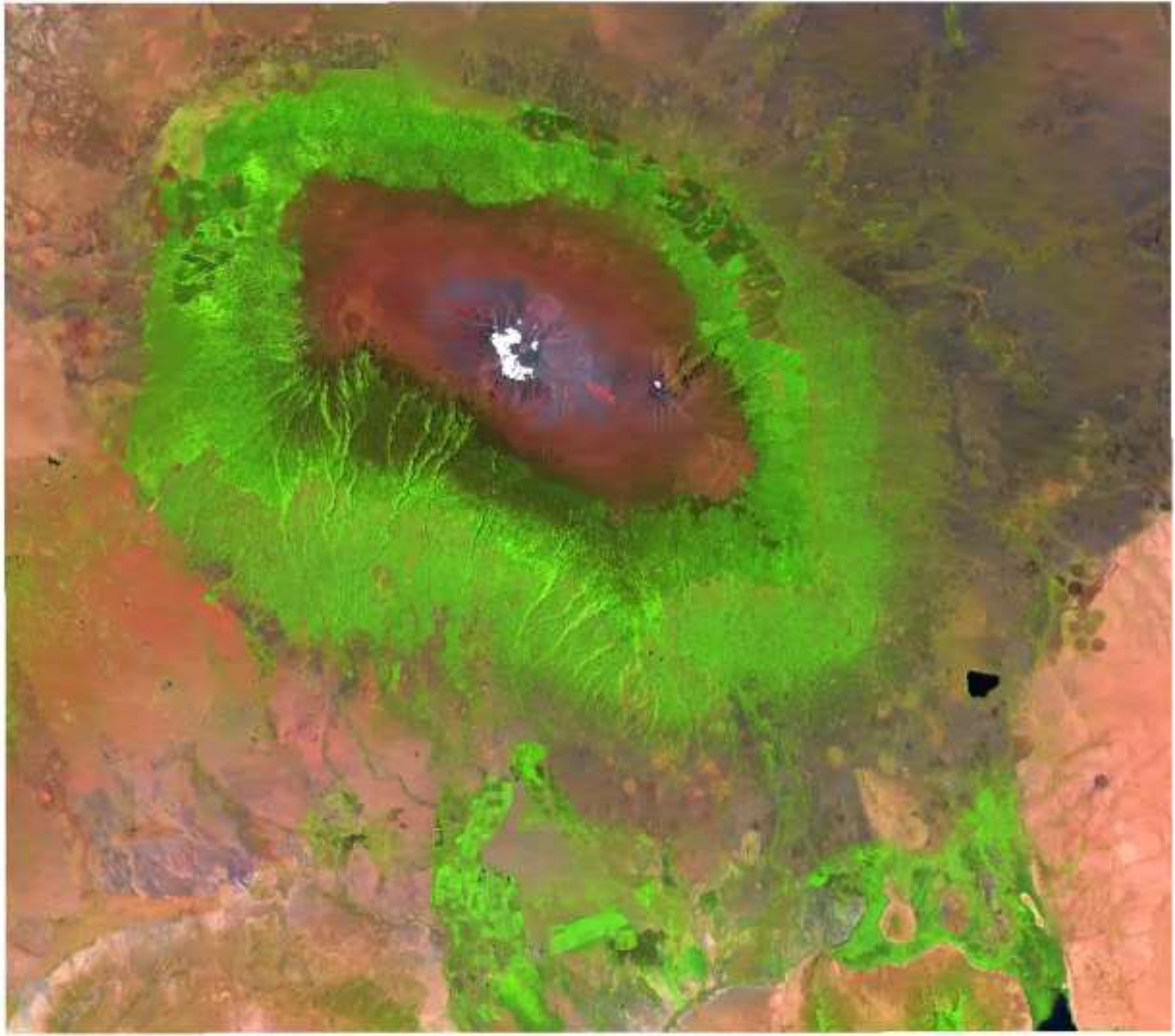
SYKE

- Temporal minimum
- Artefacts from cloud shadows visible

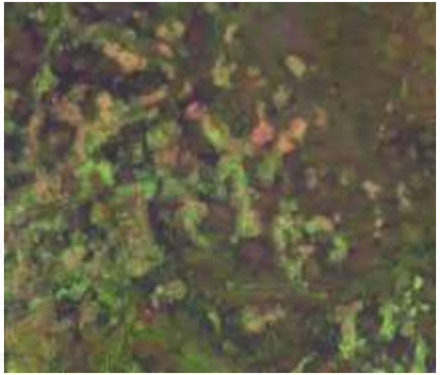


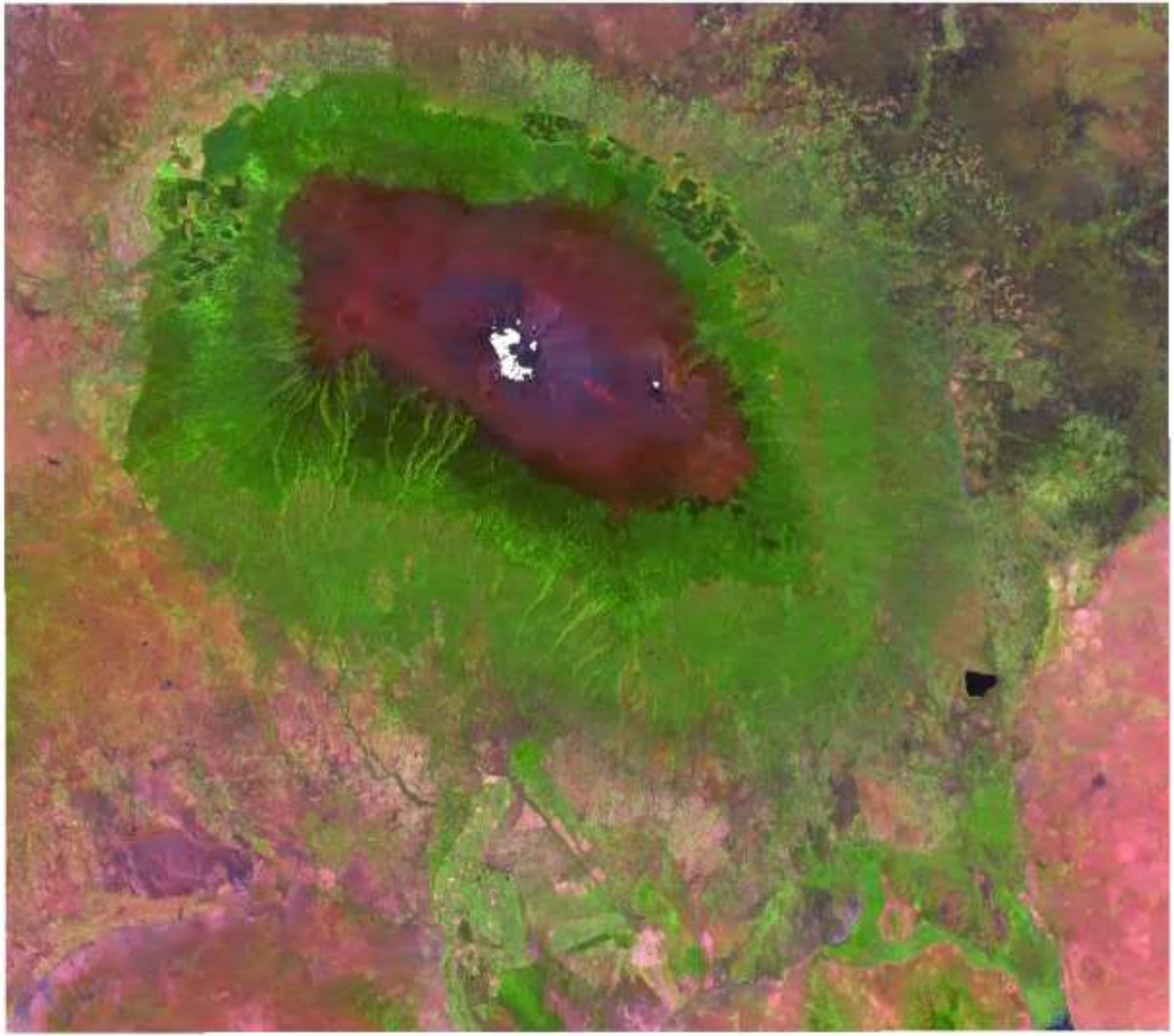
SYKE

- Temporal 0.25 quantile
- Artefacts minimized



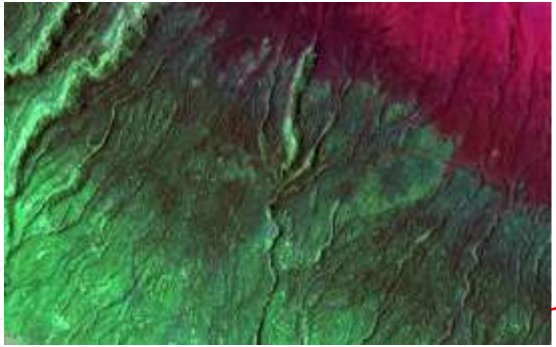
- Temporal maximum
- Cloud artefacts visible

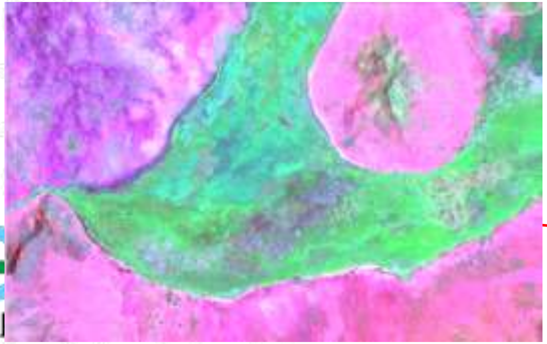


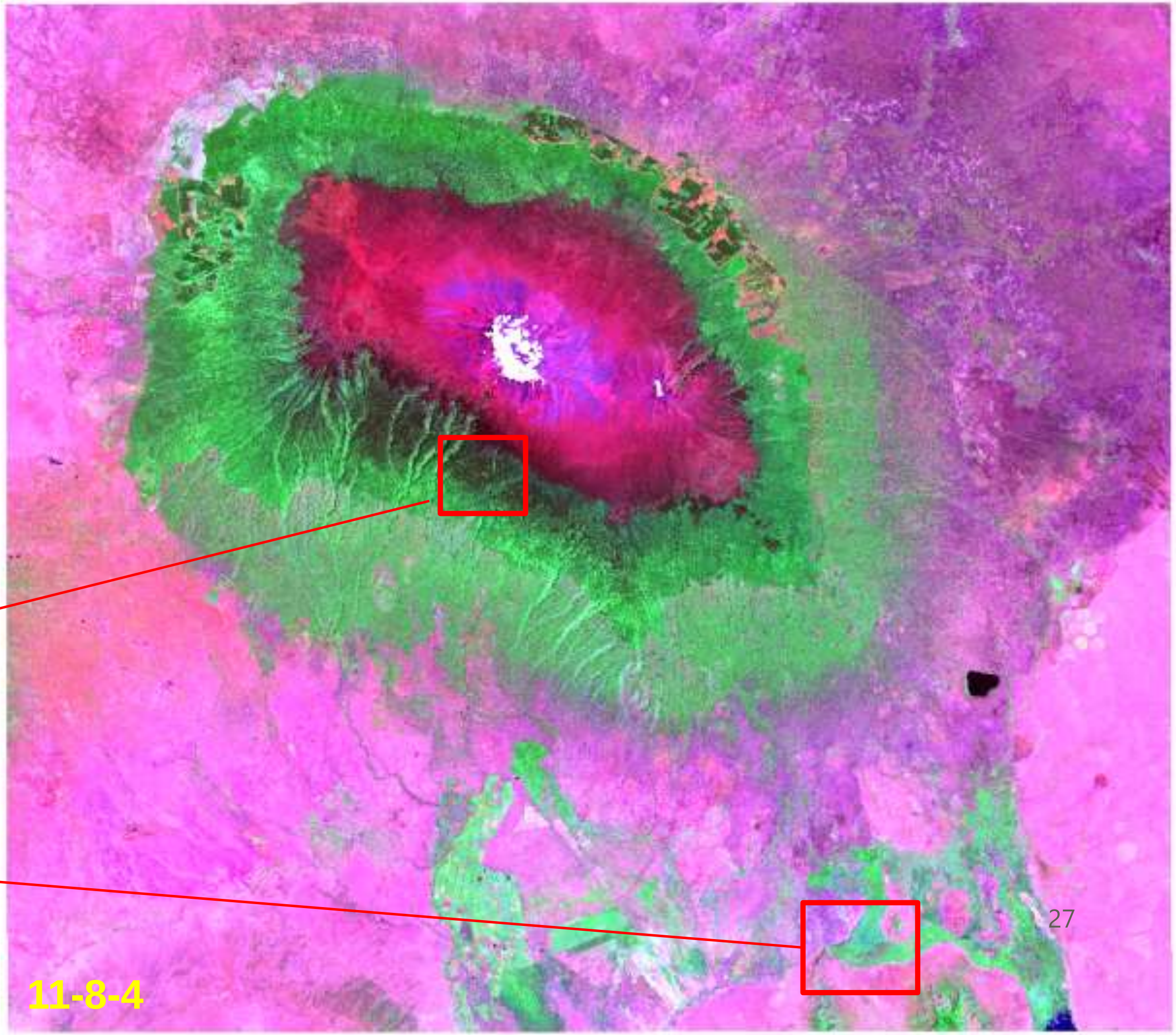


SYKE

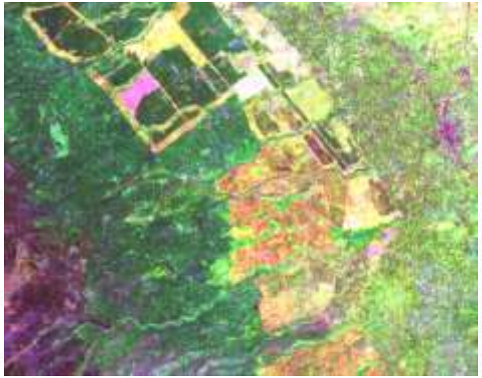
- Temporal 0.75 quantile
- Artefacts minimized

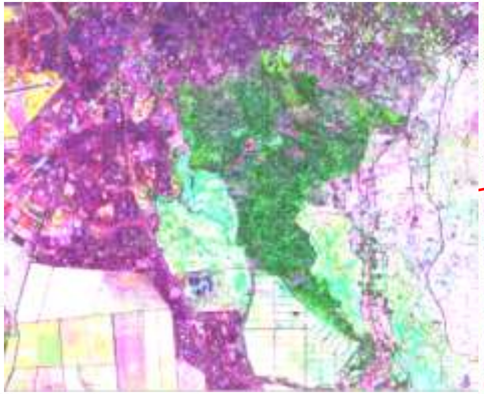


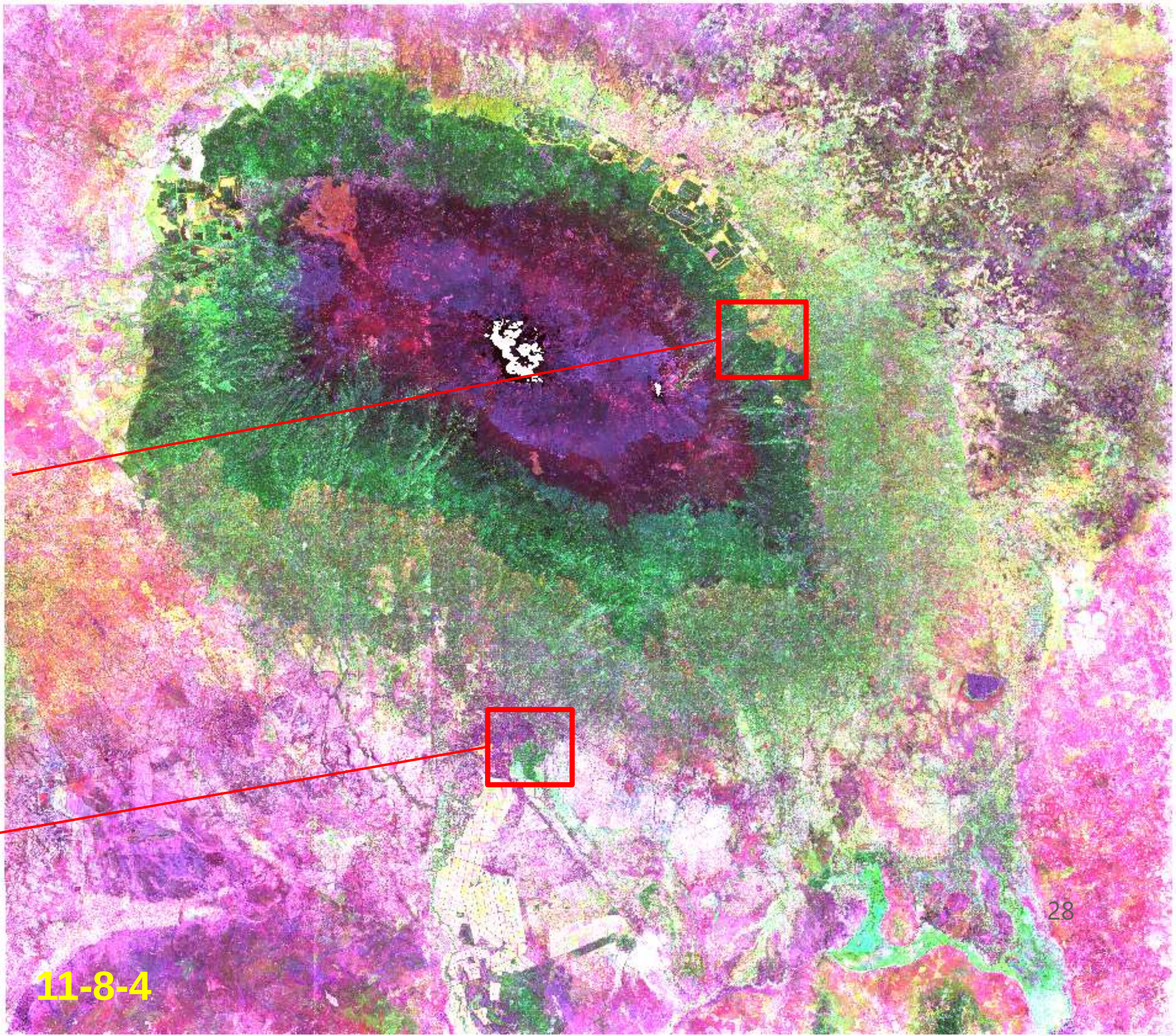




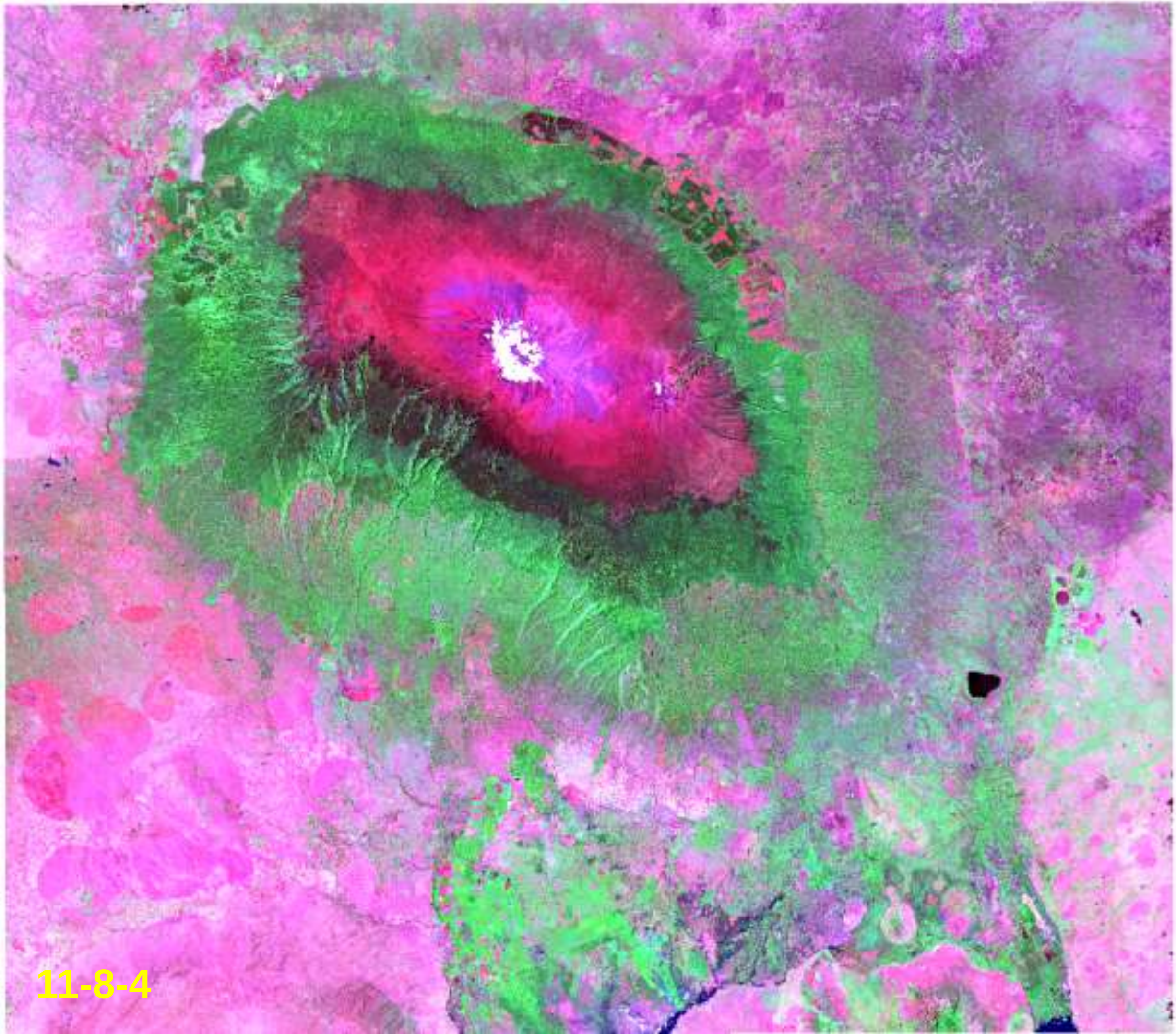
- Temporal range
- Artefacts visible (better to use interquartile range
- Reveals spectral variance







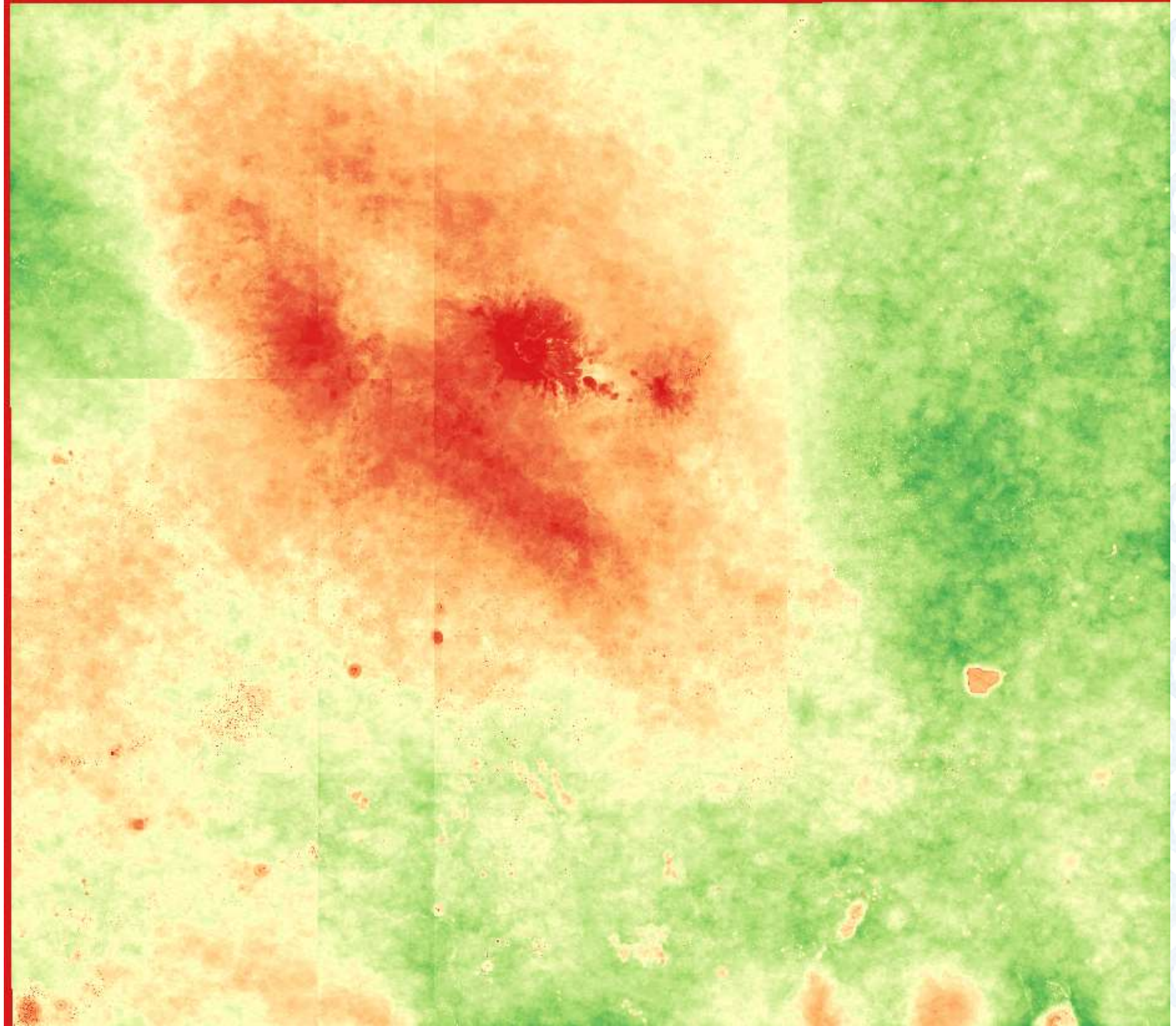
- Best Available Pixel composite
- This example is not usable due to high phenological variance and three years of observations
- Tuning of the compositing scores is also critical



- Number of cloud-free observations in compositing period

Number of cloud-free observations

- 0
- 15
- 30
- 45
- 60

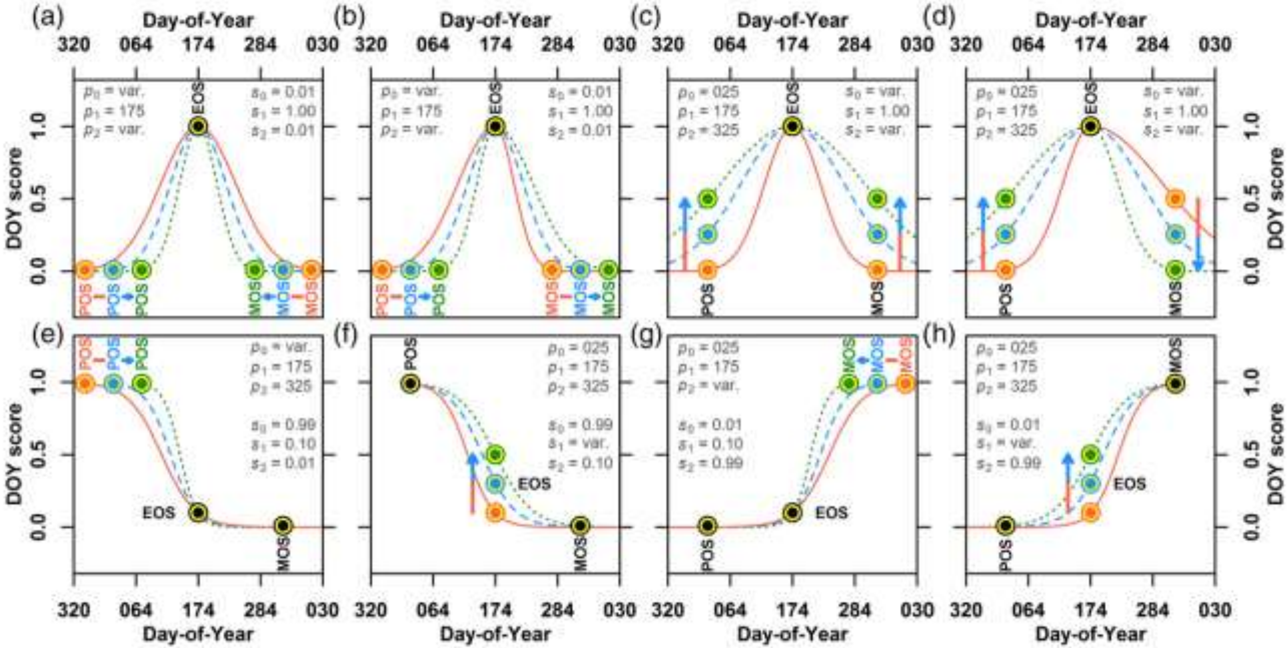


Temporal compositing scores: LSP-adaptive DOY scoring functions

Frantz et al. 2017

Gaussian, [0,1]
$S_0 < S_1 > S_2$

Logistic, [0,1]
$1 \geq s_0 > s_1 > s_2 \geq 0$
(*descending sigmoid*)
$0 \leq s_0 < s_1 < s_2 \leq 1$
(*ascending sigmoid*)



p_0, p_1, p_2 Land Surface Phenology (LSP) metrics
s_0, s_1, s_2 Function values for the LSP metrics

SYKE

Temporal compositing scores: Y-factor Y_f

D. Frantz et al. / Remote Sensing of Environment 190 (2017) 331–347

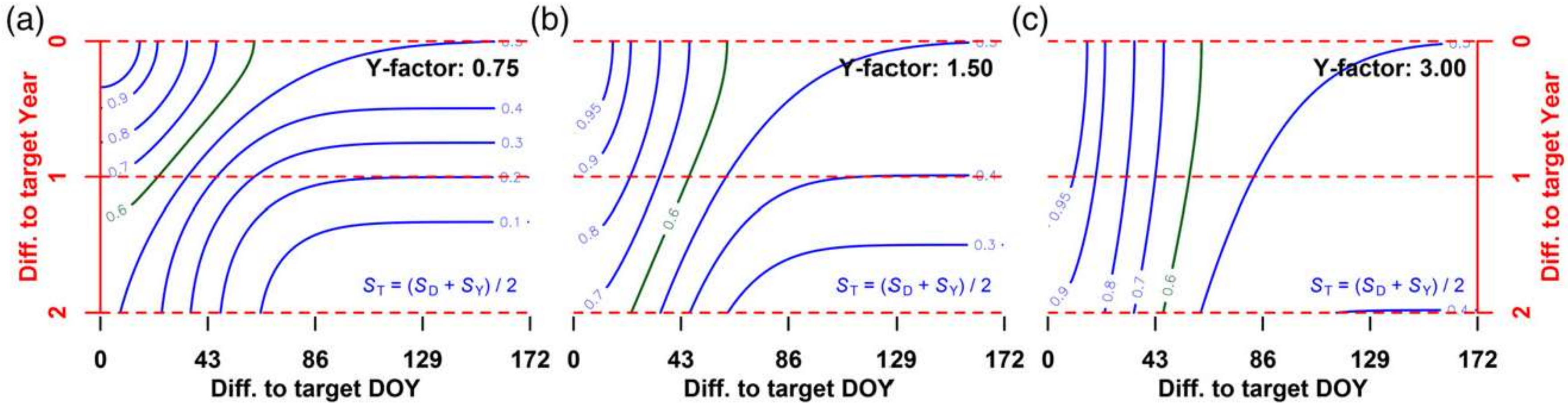


Fig. 5. Illustration of the Y-factor Y_f for an arbitrary example. The figure shows isolines of the total score S_T, drawn with a dependence on ΔY and ΔD. The right tail Gaussian type was chosen and two years around the target year are allowed; i.e. $y = 2$. The total score was computed as a combination of the Year score S_Y and the DOY. The higher Y_f, the higher is S_Y – and the smaller is the influence of ΔY on S_T.

Auxiliary compositing scores, 1/3

- **Cloud distance score**, devaluates pixels in close proximity to a cloud or cloud shadow
  - uses modified *fmask* code, with $\mathbf{d_{req}}$ parameter (distance in meters beyond which the sky is assumed to be clear)
- **Haze score**, devaluates hazy observations using *Haze Optimized Transformation* (HOT) (Zhang et al. 2002, Zhu & Woodcock 2012)

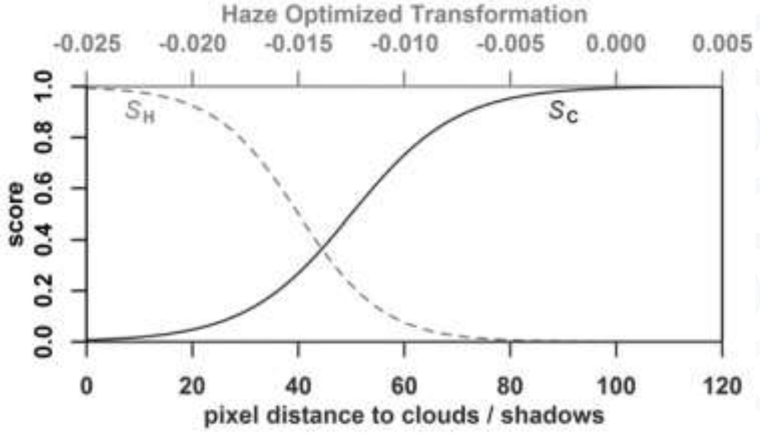


Fig. 6. Cloud distance and haze scoring functions; d_{req} was set to 100 pixels.

Auxiliary compositing scores, 2/3

- **Correlation score**, a general criterion to account for outlier-induced noise not detected previously
  - data artifacts,
  - residual misregistration,
  - undesired phenomena e.g. fires, missed clouds or shadows
- Correlates one pixel observation with other observations -> **Computationally expensive!**

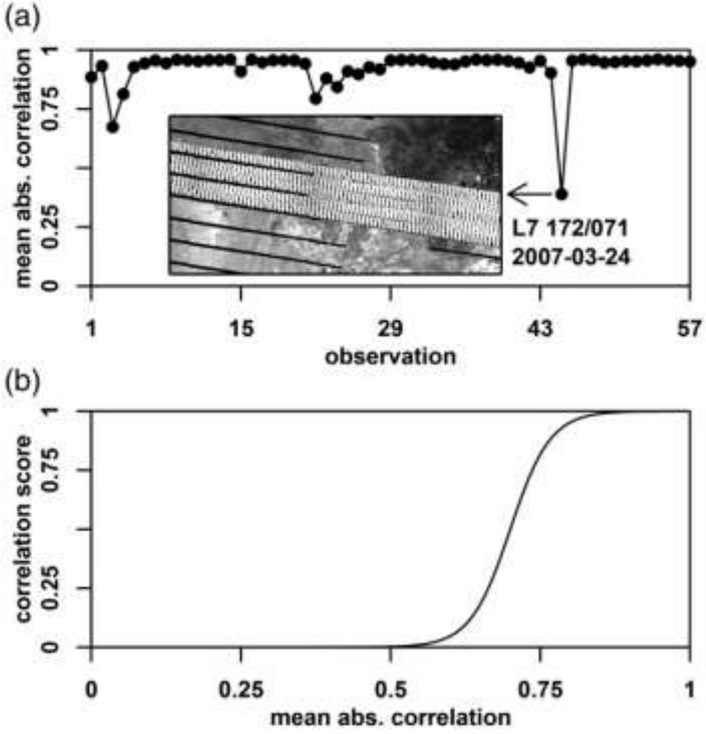


Fig. 7. Exemplary illustration of the spectral stability assessment and correlation scoring function. (*a*): mean absolute correlation derived from all spectra within the compositing period, and Landsat-7 NIR image (Path/Row 172/072) affected by sensor anomalies. (*b*): correlation scoring function.

THE NAME OF THE PRESENTER, SYKE 1.10.2019

SYKE

Auxiliary compositing scores, 3/3

- **View angle score** – devaluates off-nadir pixels on the basis of view zenith angle (e.g. for MODIS data)

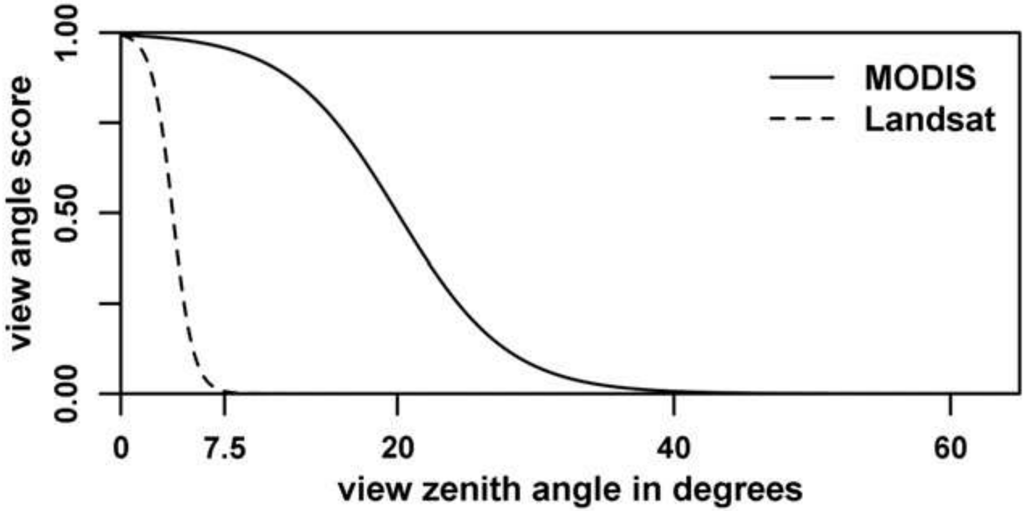


Fig. 8. View zenith angle scoring function; θ_{req} was set to 40° and 7.5° for MODIS and Landsat, respectively.

Weighting parameters W

- You can also give weights [0,1] for each compositing score S (DOY, Year, Cloud, Haze, Correl, Vzen)
- Weight = 0 disables the score entirely

FORCE TSA – Time Series Analysis

- Extracts quality-controlled time series with a number of aggregation and interpolation techniques
  - TS on spectral bands, vegetation indices, spectral mixture analysis
  - Annual Land Surface Phenology metrics
- Change and trend analyses can be done on any of the TS
- TSA needs Level 2 ARD as input!

SYKE

Phenology-based CAT-analysis

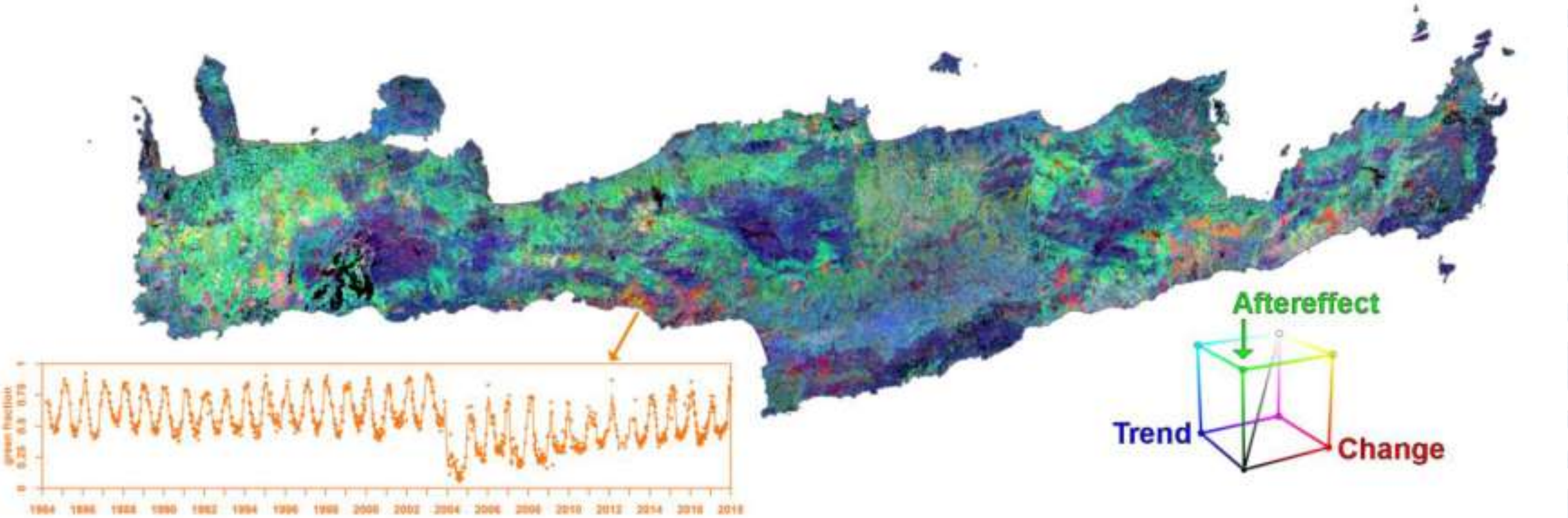


Fig. 4. Phenology-based change and trend analysis.
Change, Aftereffect, Trend transformation (CAT) showing both long-term (30+ years) gradual and abrupt changes over Crete, Greece. The CAT transform was applied to the Value of Base Level (VBL) annual time series, which was itself derived by inferring Land Surface Phenology (LSP) metrics from dense time series of green vegetation abundance derived from linear spectral mixture analysis (SMA).
[All this was done in one step using force-tsa; then mosaicked using force-mosaic]

ImproPhe – Improve spatial resolution of LSP

- Data fusion tool to developed to refine LSP metrics from MODIS using Landsat or S2

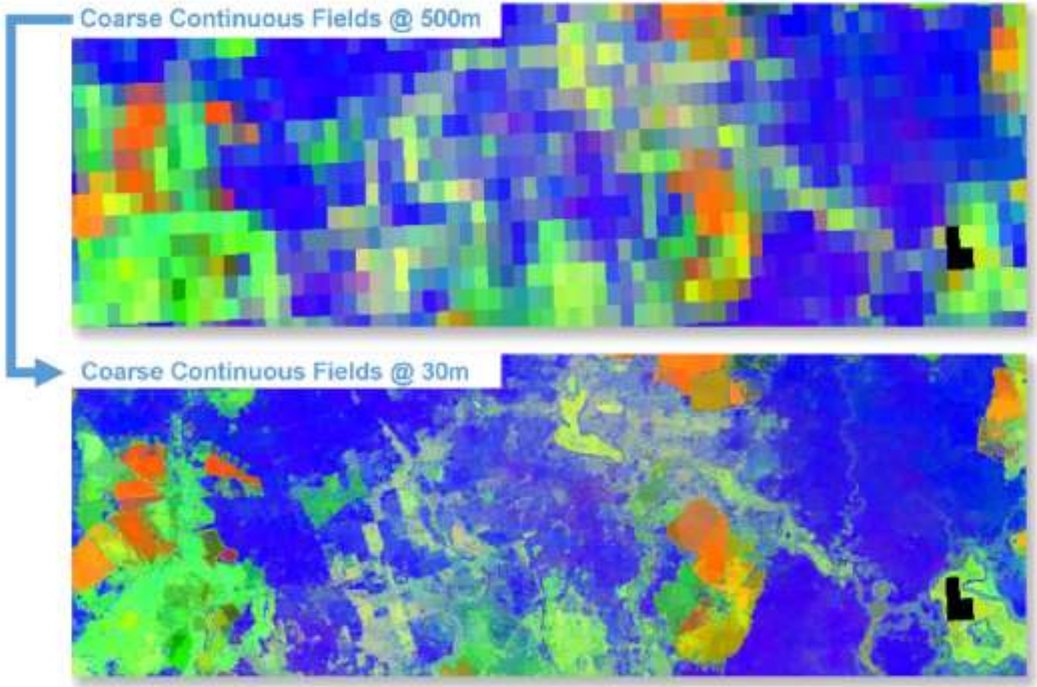


Fig. 8. Coarse resolution (500m) and ImproPhed (30m) LSP metrics.
Rate of Maximum Rise (R), Integral of Green Season (G), and Value of Early Minimum (B) phenometrics for a study site in Brandenburg, Germany. Using the ImproPhe algorithm, the LSP metrics were improved to 30m spatial resolution using Landsat and (degraded) Sentinel-2 targets.
[Data were fused using force-improphe]

ImproPhe – Improve spatial resolution of LSP

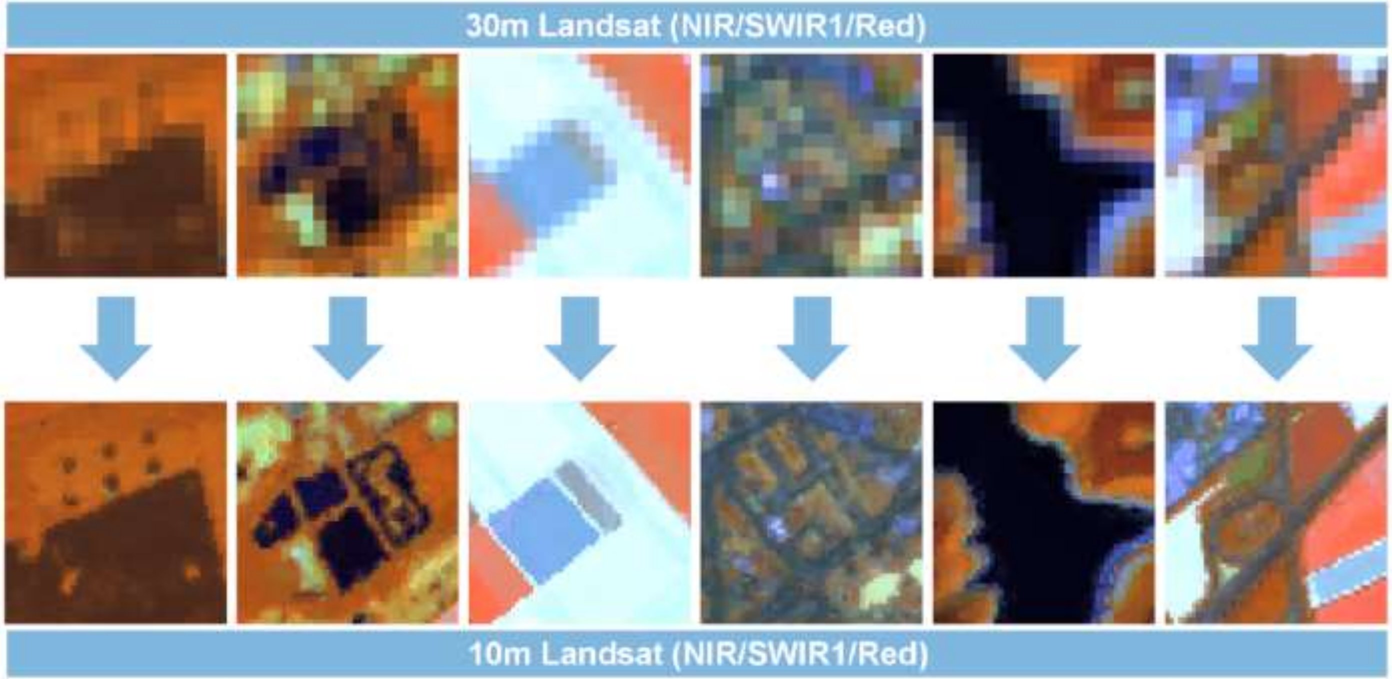


Fig. 9. 30m Landsat ARD, and ImproPhed 10m Landsat ARD.
The figure shows image subsets from North Rhine-Westphalia, Germany. Using the ImproPhe algorithm, the spatial resolution was improved to 10m using multi-temporal Sentinel-2 A/B high-res bands as prediction targets.
[Data were fused using force-l2imp]

SYKE