FORCE

Framework for Operational Radiometric Correction for Environmental Monitoring

v. 2.0

Technical User Guide

David Frantz

Humboldt-Universität zu Berlin
Geography Department
Geomatics Lab
10099 Berlin – Germany

Contact: david.frantz@geo.hu-berlin.de

Downlod: force.feut.de

Twitter: #FORCE_EO

last edit:

October 17 2018

Table of contents

List of Figures

List of Tables

I. Introduction

This user guide summarizes the technical aspects required to run the **F**ramework for **O**perational **R**adiometric **C**orrection for **E**nvironmental monitoring (**FORCE**). It will not give elaborated descriptions of methodology. For the methodological description, refer to the scientific publications below.

FORCE is intended to be an all-in-one solution for the mass-processing of selected medium-resolution satellite image archives to enable large area + time series applications. Currently supported are Landsat 4/5 TM, Landsat 7 ETM+, Landsat 8 OLI and Sentinel-2 A/B MSI. The software is capable of processing Level 1 products as obtained from the space agencies to Level 2–4 products.

The main features are

- Integration of Landsat 4–8, and Sentinel-2 A/B as Virtual Constellation.
- Data management of Landsat and Sentinel-2 Level 1 data + Download of Sentinel-2 data.
- Near-realtime (NRT) processing capability.
- Generation of Analysis Ready Data (ARD). Advanced cloud and cloud shadow detection. Quality screening. Integrated atmospheric and topographic correction: one algorithm for all sensors. Adjacency effect correction. BRDF reduction. Resolution merge of Sentinel-2 bands: 20m → 10m. Data cubing: reprojection / gridding.
- Generation of highly Analysis Ready Data (hARD). Large area. Gap free. Best Available Pixel (BAP) composites. Phenology Adaptive Composites (PAC). Spectral Temporal Metrics (STM). Ideal input for your Machine Learners!
- Generation of highly Analysis Ready Data plus (hARD+). 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).
- 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

.

Follow FORCE on social media:

ResearchGate: https://www.researchgate.net/project/FORCE-Framework-for-Operational-Radiometric-Correction-for-Environmental-monitoring

Twitter: #FORCE_EO @d__frantz @HumboldtRemSens

II. References

Frantz, D. (2017). **Generation of Higher Level Earth Observation Satellite Products for Regional Environmental Monitoring.** *Ph.D. dissertation.* Faculty of Regional and Environmental Sciences, Trier University, Trier, Germany, p. 194
http://ubt.opus.hbz-nrw.de/volltexte/2017/1046/pdf/frantz_phd.pdf

Frantz, D., Röder, A., Stellmes, M., and Hill, J. (2016). **An operational radiometric Landsat pre-processing framework for large-area time-series applications.** *IEEE Transactions on Geoscience and Remote Sensing* 54 (7): 3928-3943.
DOI: 10.1109/TGRS.2016.2530856

Frantz, D., Röder, A., Udelhoven, T., and Schmidt, M. (2015). **Enhancing the Detectability of Clouds and Their Shadows in Multitemporal Dryland Landsat Imagery: Extending Fmask.** *IEEE Geoscience and Remote Sensing Letters* 12 (6): 1242-1246.
DOI: 10.1109/lgrs.2015.2390673

Frantz, D., Haß, E., Uhl, A., Stoffels, J., and Hill, J. (2018). **Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects.** *Remote Sensing of Environment* 215, 471-481.
DOI: 10.1016/j.rse.2018.04.046

Frantz, D., Röder, A., Stellmes, M., and Hill, J. (2017). **Phenology-adaptive pixel-based compositing using optical earth observation imagery.** *Remote Sensing of Environment* 190, 331-347.
DOI: 10.1016/j.rse.2017.01.002

Frantz, D., Stellmes, M., Röder, A., Udelhoven, T., Mader, S., and Hill, J. (2016). **Improving the Spatial Resolution of Land Surface Phenology by Fusing Medium- and Coarse-Resolution Inputs.** *IEEE Transactions on Geoscience and Remote Sensing* 54 (7): 4153-4164.
DOI: 10.1109/TGRS.2016.2537929

Frantz, D., Röder, A., Stellmes, M., and Hill, J. (2015). **On the derivation of a spatially distributed aerosol climatology for its incorporation in a radiometric Landsat pre-processing framework.** *Remote Sensing Letters* 6 (8): 647-656.
DOI: 10.1080/2150704x.2015.1070314

Frantz, D., and Stellmes, M. (2018). **Water vapor database for atmospheric correction of Landsat imagery.** *PANGAEA.* Dataset available online:
DOI: 10.1594/PANGAEA.893109

Doxani, G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., and Vanhellemont, Q. (2018). **Atmospheric Correction Inter-Comparison Exercise.** *Remote Sensing* 10, 352.
DOI: 10.3390/rs10020352

A Bib TeX file with relevant methods is included with FORCE,

as is a Bib TeX file with studies that applied FORCE
(send me your published papers with DOI, they will be featured in the next version):

david.frantz@geo.hu-berlin.de

III. USAGE POLICY, DISCLAIMER, ACKNOWLEDGEMENT AND CONTACT

Proper acknowledgement, and the citation of the scientific publications are kindly requested. Co-authorship of scientific articles should be offered in case I provide substantial help in setting up, modifying or running the software, or if the generated data are an integral part of your research. Please address questions to David Frantz, Humboldt-University Berlin (david.frantz@geo.hu-berlin.de). I will try my best to answer your query within a reasonable amount of time. However, please note that I cannot guarantee user support or an immediate answer, as this is not part of my full-time job. For inquiries about the installation, please refer to section IV first; and if unsuccessful ask your administrator to set up the software before making contact. Inquiries about porting the code to non-supported operating systems, requesting a GUI, or similar requests will be ignored. Please read this user guide, as well as the scientific publications thoroughly before making any inquiries.

FORCE was developed by David Frantz at Trier University during the course of his Ph.D. thesis supervised by Prof. Dr. Joachim Hill, online available at <http://ubt.opus.hbz-nrw.de/volltexte/2017/1046/pdf/frantz_phd.pdf>, as well as in the projects ‘Southern African Science Service Centre for Climate Change and Adaptive Land Management Project’ supported by the Federal Ministry of Education and Research under contract number FKZ-01LG1201C, and ‘Sentinel4GRIPS’ supported by the Federal Ministry of Transport and Digital Infrastructure and National Aeronautics and Space Research Centre of the Federal Republic of Germany under contract number FKZ-50EW1605, with intellectual contributions from Prof. Dr. Joachim Hill, Dr. Achim Röder, Dr. Marion Stellmes, as well as code fragments from Dr. Sebastian Mader. New FORCE functionality is currently developed by David Frantz at Humboldt-University Berlin in the project ‘Near-Real Time Derivation of Land Surface Phenology using Sentinel Data: the FORCE-NRT approach’, supported by the Geo.X Research Network for Geosciences in Berlin and Potsdam. Many thanks to the geomatics team at HUB for many new ideas and methods that have made it into FORCE.

IV. Hardware requirements and Installation

Download **FORCE** from force.feut.de. FORCE is free, but registration is required.

A. *Hardware*

FORCE is intended for mass processing of medium-resolution satellite image archives. Thus, the hardware requirements are closely tied to the data volume of the envisaged project (space / time), as well as to the type of sensor (RAM / storage requirements for Sentinel-2 are higher than Landsat 4–8). Although the framework can also be used to process single (or a few) images, we generally recommend to use multi-CPU server systems. FORCE is command line software; a GUI is not provided.

It is advised to monitor RAM usage, and we recommend to decrease the number of CPUs if swapping occurs. The software installation itself is small, but the disk should be large enough for the envisaged project (commonly a couple of Terabytes); note that the software does not check for free space. Internet access is needed for some optional functionality.

B. *Operating system*

The software was developed and tested under Ubuntu 12.04 LTS, 14.04 LTS and 16.04 LTS operating systems. Note that we do not provide support for migrating the code to any other OS.

However, we were able to successfully install and run FORCE on the Windows 10 subsystem for Linux, which can be enabled in developer mode (see https://www.howtogeek.com/249966/how-to-install-and-use-the-linux-bash-shell-on-windows-10/). Please note that this procedure is in beta stage and might not be as stable as running FORCE on a real Linux system.

C. *Required software*

There are a number of required open-source dependencies which need to be installed before attempting an installation.

- GNU parallel is used for some parallelization tasks.

  We developed the code using version 20140322.
  The software can be installed with 'sudo apt-get install parallel' (Ubuntu).

  Parallel has to use --gnu mode, not --tollef. If --tollef is your default (occurred on older installations), fix this permanently by deleting the --tollef flag in /etc/parallel/config. Refer to the manpage of parallel for details.
  Parallel will display a citation request. To silence this citation notice run 'parallel --bibtex' once or use '--no-notice'.

- The GDAL API and raster utility are used for I/O and various image processing tasks.

  We developed the code using version 2.2.2.
  Note that different Ubuntu versions come with different GDAL versions.

  Ubuntu 18.04 LTS:
  The software can be installed with
  ```
  sudo apt-get install libgdal-dev gdal-bin python-gdal
  ```

  Ubuntu 16.04 LTS (and before):
  Upgrade GDAL by adding the unstable UbuntuGIS repository:
  ```
  sudo add-apt-repository ppa:ubuntugis/ubuntugis-unstable
  sudo apt-get update
  ```
  If GDAL was already installed, upgrade:
  ```
  sudo apt-get dist-upgrade
  ```
  If not, install with
  ```
  sudo apt-get install libgdal1-dev gdal-bin python-gdal
  ```

  There are known problems with earlier releases (< 1.10.0), there are no known problems with later releases.
  However, the reporting of errors and warnings differs between versions, and GDAL may report many non-critical errors to stderr (e.g. ERROR 6 – not supported, please refer to the error code description whether these are critical errors or just warnings that can be ignored). Please note that GDAL is a very dynamic development, therefore it is hard to develop applications that cover all cases and exceptions in all possible GDAL versions and builds.

- The GSL library is used for optimization purposes.

  We developed the code using version 1.15.

The software can be installed with ‘sudo apt-get install libgsl0-dev’ (Ubuntu).

There are no known problems.

- The cURL library is used to download MODIS water vapor data.

We developed the code using version 7.22.0.
The software can be installed with ‘sudo apt-get install curl’ (Ubuntu).

There are no known problems.

- unzip is used to extract Sentinel-2 data.

We developed the code using version 6.
The software can be installed with ‘sudo apt-get install unzip’ (Ubuntu).

There are no known problems.

- lockfile-progs is used to place a temporary lock on file queues

The utility is already included in some distributions.
The software can be installed with ‘sudo apt-get install lockfile-progs’ (Ubuntu).

There is a known problem with CIFS mounted network drives. You may get a lot of warnings like ‘lockfile creation failed: exceeded maximum number of lock attempts’. You can ignore these warnings; they are no fatal errors. But you might want to inspect the file queue after Level 2 processing, as there is a minor possibility that there were some conflicts due to parallel write attempts: a few images might not have been switched from QUEUED to DONE status.

- rename is used to rename files.

The tool is missing in new Ubuntu distributions (Ubuntu > 17.10).
The software can be installed with ‘sudo apt-get install rename’ (Ubuntu).

There are no known problems.

D. Optional software

Some FORCE functionality is dependent on additional open-source software. In order to allow a baseline installation (in consistency with FORCE v. ≤ 1.1), these dependencies are optional and are not required for the default installation. If you want to enable this functionality, you need to install following software, and follow the optional steps in section F.

- SPLITS (Spline analysis of Time Series), developed by Dr. Sebastian Mader, Trier University, is a standalone software/API, which can be used to fit splines to dense time series in order to interpolate time series and to derive Land Surface Phenology.

Some SPLITS-dependent features are implemented in *force-tsa* (see section VII.E). You only need to install SPLITS if you want to use these features; all other functionality of *force-tsa* can be used without installing SPLITS.

SPLITS is distributed under the terms of the GNU General Public License, and can be downloaded from http://sebastian-mader.net/splits/. SPLITS is itself dependent on GDAL (required for FORCE, see last section), Armadillo and FLTK.

For installation instructions, see http://sebastian-mader.net/splits/, or follow the following lines (Ubuntu). Make sure to change the path to your own home directory:

```
sudo apt-get install libarmadillo-dev
sudo apt-get install libfltk1.3-dev

mkdir /home/MYHOME/src
mkdir /home/MYHOME/splits

cd /home/MYHOME/src
wget http://sebastian-mader.net/wp-content/uploads/2017/11/splits-1.9.tar.gz
tar -xzf splits-1.9.tar.gz
cd splits-1.9

./configure --prefix=/home/MYHOME/splits CPPFLAGS="-I /usr/include/gdal" CXXFLAGS=-fpermissive
make
```

make install
make clean

There are no known problems, except that you need to install SPLITS v. 1.9.

E. Installation

FORCE needs to be compiled. Administrator rights are not necessarily required, unless you want to install to the system-wide search path. It is advised to include the installation directory (BINDIR) in $PATH.

1) Extract the tarball; e.g. ‘tar –xzf force_v1.1_beta.tar.gz’; the output path can be defined with the –C switch.

2) Edit the Makefile

   BINDIR is the directory used for installing the software; e.g. ~/bin

   LIBDIR is a directory for compiling a dynamic library used within FORCE; e.g. ~/lib

   Both directories need to exist.

   GDAL names the directories, where GDAL header files and library are installed.
   You can use ‘gdal-config --cflags’ and ‘gdal-config --libs’ to retrieve the directories.

   Most compiling errors are related to GDAL, thus make sure that the compiler finds the GDAL installation.

3) Navigate to the directory containing the Makefile and type ‘make’.

F. Installation with optional software

Install with SPLITS: follow these steps before compiling FORCE, i.e. step 3 in section E:

1) Install SPLITS (see section D)

2) Edit src/force.h

   In the compiler options section, uncomment the SPLITS definition by removing ‘//’ before ‘#define SPLITS’.

3) Edit the Makefile

   SPLITS names the directories, where SPLITS header files and library are installed. Example:
   SPLITS=-I/home/MYHOME/splits/include/splits/ -L/home/MYHOME/splits/lib

   LDFLAGS2 links FORCE with the optional software libraries. Use:
   LDFLAGS2=-lsplits -larmadillo

V. Version History

A. *FORCE v. 1.0 beta*

- Release: 07.10.2017
- First public release.

B. *FORCE v. 1.1 beta*

- Release: 11.02.2018
- FORCE L1AS
  - Download and management support for Sentinel-2B, higher level support was already implemented in v. 1.0 beta.
  - The Sentinel-2 downloader/archiver has 4 new mandatory arguments/filters: sensing-start-time sensing-end-time min-cloud-cover max-cloud-cover.
  - Fixed a bug when importing Landsat images with white-space characters in the directory name; now, this throws an error.
- FORCE L2PS
  - Reduced memory requirements for holding TOA reflectance internally by ~0.72GB (Landsat 4-7), ~0.95GB (Landsat 8), ~2.92GB (Sentinel-2)
  - Reduced memory requirements when correcting for adjacency effects by ~1.20GB (Landsat 4-8), ~2.24GB (Sentinel-2)
  - Fixed a bug in the BRDF correction of the Landsat 8 SWIR-2 band.
  - BRDF correction is now available for Sentinel-2.
  - Relative Spectral Response Function of Sentinel-2B is now included.
  - Relative Spectral Response Function of Landsat 4 TM was updated.
  - Improved gaseous vapor transmittance calculations. The water vapor absorption spectrum was updated to the HITRAN 2016 database (obtained from HITRAN on the Web, http://hitran.iao.ru/) The computation of water vapor absorption and water vapor and ozone transmittance was corrected: a numerical error in calculating transmittance resulted in increased SWIR-1 reflectance in Landsat-TM data.
  - The spectral resolution of all tables was reduced from 0.1 nm to 1 nm.
  - Added bandnames to all output products.
  - Cloud shadow matching for Sentinel-2 images is now performed on the 20m resolution, which significantly speeds up the work, especially for poorly illuminated winter images where the base height is iterated in small vertical increments.
- FORCE L3PS
  - Added new parameter OFF_SEASON, which allows the compositing algorithm to select clear-sky observations that are not within the prime season of interest. This may be advantageous if gap-free images are needed. This parameter also affects the temporal-spectral metrics. If enabled, the statistics will be derived over the complete year, instead of the prime season of interest.
  - Added new parameters OUTPUT_Q25, OUTPUT_Q50, OUTPUT_Q75 and OUTPUT_IQR that allow the computation of 25%/50%/75% quantiles and interquartile range, defined as Q75 - Q25.
  - Added new parameter LSP_START, which allows you to input LSP data with a custom starting point. Previously, the LSP data had to be provided as YEAR*365+DOY. The new parameter specifies an offset as Y*365+DOY. Thus, if the LSP are provided as before, LSP_START = 1. If the LSP data would be provided relative to January 1 2000, LSP_START = 730001, i.e. 2000*365+1.
  - Fixed a bug, which resulted in no data when using the phenology-adaptive compositing with only one year of LSP data.
  - Removed unused parameter LSP_NUM_YEAR.
  - Added new program *force-level3-mosaic*, which generates virtual mosaics from tiled Level 3 products in Virtual Dataset format. Due to performance considerations, it is advised to enable tiled output in *force-level3*, and if necessary, use this new tool to create mosaics. *force--quicklook-level3* was updated to generate quicklooks from the vrt mosaics.
  - Added bandnames to all output products.
  - Negative tile numbers are now allowed.
- FORCE TSA
  - Added product id as metadata for time series stack output.

    - Parallel processing can be used. New parameter was added to parameter file. Computation of TC-DI is currently not parallelized.
    - Added bandnames to all output products.
    - Negative tile numbers are now allowed.

- FORCE WVDB

    - Collection 5 geometa tables were removed from LAADS' FTP server, leaving *force-lut-modis* nonfunctional. The code was updated to make use of collection 6.
    - Discovered a randomly occurring bug, seemingly related to a data race in localtime(): the function timer was set to tomorrow, resulting in a never-ending while loop. Measures were implemented to avoid this data race.

- FORCE AUX

    - Added a new program *force*, which prints the version number and a short disclaimer.
    - Fixed a bug of missing resolution in *force-qai-inflate* outputs.
    - *force-tabulate-grid* was completely overworked. Instead of a table with coordinates, a shapefile is directly generated, which let's you easily relate to the processing grid.
    - Added bandnames to all output products (*force-qai-inflate*).

C. *FORCE v. 2.0*

- Release: 02.10.2018

- General changes

    - Follow FORCE on social media. ResearchGate: https://www.researchgate.net/project/FORCE-Framework-for-Operational-Radiometric-Correction-for-Environmental-monitoring. Twitter: #FORCE_EO.
    - There are many new features throughout all FORCE modules, and there are entirely new modules. The new features generally come with new parameters in the parameter files, which you need to add to your existing parameter files (from old FORCE versions). If you are seeing an error like this, simply add the new parameters (have a look at the user guide for details and possible values): 'OUTPUT_FORMAT is not specified. Unable to read parameter file!'
    - All products can now be output as ENVI Standard or GeoTiff. An ENVI header is written for both options. The first option produces uncompressed flat binary files. The GeoTiff come with LZW-compression using horizontal differencing. A new parameter OUTPUT_FORMAT was added to all parameter files. All FORCE modules can digest both formats.
    - All FORCE modules are more verbose and will show some progress bar.
    - The writing of all output is now accompanied by locking the file (even if it does not exist yet). This prevents possible data races when writing one and the same output file from two threads. Note that there is the possibility that the lockfile cannot be generated, in which case no output can be written. If this happens, please refer to the FAQ section in the user guide to resolve this issue. This generally occurs if you have aborted FORCE before (lockfile was not deleted correctly) or if too much I/O is causing a timeout of the lock attempt.
    - Directories are now created with 775 permissions instead of 700. Note that your umask settings might reduce these permissions.
    - If there is a reading error for any file, the error message includes the filename.
    - There are more metadata written to all outputs, e.g. the parameterization, information about your system, processing time, FORCE version, band names etc.
    - A bibtex file including FORCE citations is now included in the software package.
    - A bibtex file including studies that used FORCE, is now included in the sofware package. If you publish anything enabled or related to FORCE, please send me your published papers (pdf+DOI). They will be included in the bibtext file of the next version: david.frantz@geo.hu-berlin.de, #FORCE_EO.
    - Recognizing the different user needs, quality screening for all higher level FORCE modules (above Level 2) is now in full user control. The parameter SCREEN_QAI was added to all parameter files. Have a look at the user guide for all possible keywords and quality filters. Please make proper use of the quality information and never analyze any data without.
    - The sensor list (SENSORS = ...) in the parameter file of all higher-level FORCE modules (above Level 2) has two additional key words: sen2a and sen2b (all lowercase) refer to the original Sentinel-2A/B 10m bands only. Note that FORCE matches the Landsat NIR band with the original Sentinel-2A/B 20m NIR band; therefore, using "SENSORS = LND08 sen2a" will only return the VIS bands. Have a look at the matching table FORCE uses for combining several sensors, and read the next change entry.
    - The filenames of the higher-level FORCE outputs (above Level 2) have changed slightly. Before, the tag "MULTI" was a placeholder for the sensor. Now, it is one of the following and merely refers to a named set of spectral bands instead of individual sensors: 1) "LNDLG" refers to the 6 Landsat legacy bands (B/G/R/NIR/SW1/SW2). It occurs if

you select one or several Landsat sensors (even if you additionally use SEN2A and/or SEN2B). "SEN2L" refers to the 10 Sentinel-2 land surface bands (B/G/R/RE1/RE2/RE3/PNIR/NIR/SW1/SW2). It occurs if you only select SEN2A and/or SEN2B. "SEN2H" refers to the 4 high-resolution Sentinel-2 bands (B/G/R/PNIR). It occurs if you only select sen2a and/or sen2b. "R-G-B" refers to visual bands only. It occurs if you select one or several Landsat sensors plus sen2a and/or sen2b. See the matching table above for details.

- The spatial data cube definition is appended to each gridded data cube. The file contains (1) projection as WKT string, (2) origin of the tile system as geographic Longitude, (3) origin of the tile system as geographic Latitude, (4) origin of the tile system as projected X-coordinate, (5) origin of the tile system as projected Y-coordinate, and (6) width of the tiles in projection units.

- FORCE L1AS

  - force-level1-sentinel2 shows the filename of the current download, progress bar, downloaded size, download speed and eta per download. Note that the progress bar and eta are produced by wget. Afaik, there is no way to change the format.
  - The L1AS programs have a new mandatory argument. The file name of the file queue needs to be given. This enables you to manage several project in one data pool. If the file queue exists, new files will be appended at the end. If not, the file will be created.

- FORCE L2PS

  - The GeoTiff driver is more picky with the output projection. Make sure to validate your WKT string using "gdalsrsinfo -v 'YOUR_WKT_STRING'". I have improved the definition of the default projection in force-parameter-level2 to work with GeoTiff.
  - The number of CPUs can be re-adjusted while(!) force-level2 is running. A file 'cpu-$TIME' is temporarily created in DIR_TEMP. This file can be modified. Note that the effect is not immediate, as the load is only adjusted after one of the running jobs (images) is finished.
  - Added a custom delay for starting new jobs in force-level2. This prevents a huge I/O-related bottleneck at the beginning of processing. As a result, it may take a while (depending on how many cores you are using) until force-level2 runs at full efficiency. An additional parameter needs to be given for force-level2 (delay in seconds).
  - force-level2 shows eta (this will only show reasonable numbers after acouple of jobs are finsihed). The estimate is based on the runtime of finished jobs, so the first estimate will only be shown when the first job has finished. Note that the progress bar and eta are produced by GNU parallel. Afaik, there is no way to change the format; sorry for the eta in seconds...
  - The processing time is written to the logfile.
  - There is a slight problem, which seemingly occurs on CIFS-mounted windows shares. Attempting to lock the file queues will fail and will throw a warning. This is not a fatal error. Please see the FAQ in the user guide if this happens on your system. I didn't find a solution yet.
  - Hopefully fixed a bug that changed file permissions of the file queue to read-only (this is related to "sed -i" and seems to occur on some Linux systems only).
  - Changed behaviour of the subzero QAI flag: Subzero = any band < 0.0. If reflectance < -1.0, the pixel is set to nodata.
  - Changed behaviour of the surface reflectance output when saturation is detected. Previously, reflectance was capped at 1.0. Now, values > 1.0 and < 2.0 are allowed, but still flagged as saturated. Pixels with any reflectance > 2.0 are set to nodata.
  - Improved the performance of the shadow probability computation in the cloud/shadow detection module. The shadow probability is only computed for pixels that are close enough to a cloud. A maximum cloud height of 12000m + max. sun zenith of scene is assumed; sun azimuth is not taken into account so far.
  - Snow mask switched on for Sentinel-2. Increased NDSI threshold to reduce comission because of the missing thermal band, included a SWIR2 threshold to counterbalance.
  - Reduced cloud/cloud shadow/snow/water ambiguity. Now, a pixel cannot be cloud and snow, or water and snow at the same time. Water takes precedence over snow. Snow takes precedence over cloud. Note that snow and cold clouds can be easily confused. No shadow is matched for snow pixels (will increase error of omission to a certain degree). Opaque clouds, and buffered clouds take precedence over cirrus clouds. Clouds take precedence over cloud shadows.
  - Decreased the aggresivity of the cirrus mask by incorporating a blue test. Cirri are not buffered anymore.
  - Reduced cloud shadow comission errors by incorporating the shadow probability in the computation of the shadow match similarity. This will favor stonger shadows and shadows won't simply snap to all dark features (like coniferous forest).
  - Snow coverage in percent is written to the logfile.
  - Implemented more efficient method to interpolate coarse resolution atmospheric parameters, + elevation-dependent parameters are computed for smaller elevation increments -> resulting atmospheric variables like AOD look much smoother.

- Fixed a bug in the geometric module, which appeared when warping and tiling images. The result was a possible subpixel misalignment as the images were simply warped into the target coordinate system without aligning the new extent with the tiling grid.
- The conversion from DN to physical units is now based on the reflectance calibration factors for Landsat as suggested by Kurt Thome @ACIX. As the ESUN spectrum is not used anymore, the remove-then-add calibration employed for Sentinel-2 was removed, too.
- There are three different resolution merge options available to increase the spatial resolution of the 20m Sentinel-2 bands to 10m. The default algorithm has changed from REGRESSION to IMPROPHE. Other options are STARFM and NONE. See user guide for more details.
- The use of the DODB was deprecated. The parameters were removed from the parameter files. This feature has never made it into an official FORCE release; before the release of FORCE v. 1.0, measures were implemented to make the AOD estimation more reliable; therefore results were fine without using it. Given the extra processing step (non-linearity in processing) and used disk space, I decided to completely remove this functionality.
- Fixed a potential data race when writing one and the same output file from two threads. Now, a lockfile is created to prevent this. Parallel writes can happen as FORCE implements a Level 2 compositing to reduce data redundancy, i.e. if two different Level 1 products cover the same Level 2 tile. By doing so, the compositing method was changed too. Before, it was simply overwriting with the latest data. Now, the existing file is read, then the average between the old and new data is computed, and the file is written completely again. This method produces more smooth transitions between Level 1 footprints. Note that the old overwrite with latest data method is still used for the QAI layer.
- It is now possible to process images without a DEM, although this is not recommended. If doing so, the DEM is assumed to be flat @ z=0m. Cloud detection and atmospheric correction both benefit from using a DEM. If topographic correction is to be used, L2PS will exit with an error. To disable, use FILE_DEM = NULL, and DO_TOPO = FALSE.
- Added wavelength domain tags to bandnames for BOA/TOA products.
- Data cover, cloud cover (opaque, buffered and cirrus), cloud shadow cover, snow cover and water cover is written to the metadata.
- force-quicklook-level2 has an additional argument, which defines the stretch (max vis value in linear stretch). Before, it was fixed to 1000 (10% reflectance), which is inappropriate in bright landscapes.

- FORCE L3PS

  - If the target resolution is lower than the Level 2 data, the spatial aggregation can be performed using an approximated Point Spread Function. Note that this option severely affects input time. A new parameter REDUCE_PSF was added to the parameter file.
  - force-level3 is more verbose, shows progress, eta, relative time spent for input/computing/output operations (note: it is not pure I/O time, it is rather a measurement of all input/output related things, e.g. the CPU time spent with REDUCE_PSF would count as input), and compositing statistics at the end. The eta is based on the runtime of finished tiles, so the first estimate will only be shown after the first tile has finished.
  - BAP composites look mostly crappy over water because it is a very variable surface type. Therefore, the compositing method is switched over permanent water, i.e. if > 90% of obs are water-flagged. Over water, minimum SWIR2 compositing is used, while enforcing that NIR reflectance is greater than SWIR2. Will only be used for sensors with NIR and SWIR2 band.
  - force-level3 can use Level 2 ImproPhe data (will use original data, i.e. BOA/TOA, if there is no IMP product for any specific date), generated by the new tool force-level2-improphe (see section FORCE L2IMP below). A new parameter USE_IMPROPHE was added to the parameter file.
  - Added wavelength domain tags to bandnames for reflectance products.
  - force-quicklook-level3 has an additional argument, which defines the stretch (max vis value in linear stretch). Before, it was fixed to 1000 (10% reflectance), which is inappropriate in bright landscapes.
  - force-level3-mosaic was deprectared, and replaced by the new force-mosaic, which can mosaic all FORCE-generated, tiled outputs.

- FORCE TSA

  - FORCE TSA has received a major update and many new features. There are many new parameters that you need to include in your existing parameter files.
  - If the target resolution is lower than the Level 2 data, the spatial aggregation can be performed using an approximated Point Spread Function. Note that this option severely affects input time. A new parameter REDUCE_PSF was added to the parameter file.
  - force-tsa is more verbose, shows progress, eta, relative time spent for input/computing/output operations (note: it is not pure I/O time, it is rather a measurement of all input/output related things, e.g. the CPU time spent with REDUCE_PSF would count as input). The eta is based on the runtime of finished tiles, so the first estimate will only be shown after the first tile has finished.
  - force-tsa can use Level 2 ImproPhe data (will use original data, i.e. BOA/TOA, if there is no IMP product for any specific date), generated by the new tool force-level2-improphe (see section FORCE L2IMP below). A new parameter USE_IMPROPHE was added to the parameter file.

    - Only bands that are needed to compute the requested index are read from disc. This decreases time spent for reading data. As an example, only red and NIR are read for calculating NDVI time series.
    - Fixed a bug that caused a core dump when outputting the TSS product with more than 1169 images.
    - Instead of a 'simple' spectral index or band, linear spectral unmixing can be used. Use INDEX = SMA. Several parameters were added to the parameter file. The endmember file is specified with FILE_ENDMEM. Only the abundance image related to one endmember is retained (this will likely change in future versions). The SMA can be Sum-to-One constrained using the parameter SMA_SUM_TO_ONE. The SMA can be Non-Negativity constrained using the parameter SMA_NON_NEG. The fractions can be shade normalized, using the parameter SMA_SHD_NORM. The shade spectrum (photogrammetric zero or measured shade) need to go into the last column of the endmember file. The RMSE time series of the SMA can be output using the parameter OUTPUT_RMS.
    - Red Edge bands (RE1, RE2, RE3) and the broad NIR band (BNIR) of Sentinel-2 can be used as INDEX.
    - Implemented Disturbance Index without rescaling, i.e. no spatially tuned z-transformation. INDEX = TC-DI0.
    - Time series can be interpolated. Several parameters were added to the parameter file. The interpolation method is set with INTERPOLATE. Currently implemented are NONE (no interpolation), LINEAR (linear interpolation), MOVING (running mean interpolation / smoothing), and RBF (Radial Basis Function interpolation / smoothing). The maximum temporal distance (i.e. filter width) for the MOVING filter, is set with MOVING_MAX (in days). Several kernels can be used for the RBF filters using RBF_SIGMA, the sigma(s) are given in days. The kernel strengths are adapted by weighting with actual data availability within the time covered by each kernel. The parameter RBF_CUTOFF specifies a cutoff value for temporally truncating the kernels, such that e.g. 95% of the Gaussian bell is retained. The time step for the interpolation is set with INT_DAY (in days). The interpolated time series can be output with OUTPUT_TSI.
    - Land Surface Phenology (LSP) can be derived from dense timeseries. The Spline Analysis of Time Series (SPLITS) API was incorporated for this purpose. FORCE can be compiled with and without SPLITS, in the latter case, the new functionality won't be available. See the user guide for instructions on how to install with SPLITS. A number of new parameters were added to the parameter file, and new products can be output. LSP extraction is triggered by using LSP as folding option (FOLD = LSP), or when OUTPUT_LSP = TRUE. The number of segments for fitting the splines needs to be specified using the LSP_N_SEGMENT parameter. To derive LSP metrics for a given year, some data from the previous and next year need to be included (LSP_DOY_PREV_YEAR, LSP_DOY_NEXT_YEAR). The user can select if a Northern hemisphere, Southern hemisphere or mixed phenology is expected (LSP_HEMISPHERE). In total, 26 LSP metrics will be derived for each year.
    - The change and trend (TRD/CAT) products can be computed on the annual LSP metrics. This will produce 26 change / trend products.

- FORCE CSO

    - New module FORCE Clear Sky Observations (FORCE CSO) is intended for data availability mining. For given time steps (e.g. 3 months), per-pixel statistics about data availability are calculated, i.e. number of CSOs, and average (standard deviation, min, max, etc.) days between consecutive CSOs. FORCE CSO includes the programs force-parameter-cso to generate an empty parameter file, and force-cso to do the data mining.

- FORCE ImproPhe

    - New module FORCE Improving the Spatial Resolution of Land Surface Phenology (FORCE ImproPhe) is intended to to increase the spatial resolution of coarse continuous fields. It was originally developed to refine Land Surface Phenology metrics derived from MODIS, using sparse Landsat data as spectral and multi-temporal targets for data fusion. Regarding phenology, it can be used to obtain a Landsat-like phenology even in areas / during times when Landsat data alone is insufficient (in terms of temporal density). FORCE permits the use of Landsat and/or Sentinel-2 data as target datasets for the improPhement. ImproPhe can also be applied to other coarse resolution data. FORCE ImproPhe includes the programs force-parameter-improphe to generate an empty parameter file, and force-improphe to perform the data fusion.

- FORCE L2IMP

    - New module FORCE Level 2 ImproPhe (FORCE L2IMP) is intended to increase the spatial resolution of lower resolution Level 2 ARD using higher resolution Level 2 ARD, e.g. to improve the spatial resolution of 30m Landsat imagery to 10m using Sentinel-2 data as targets. This only works for years where both data sources exist. The data fusion is performed with the ImproPhe algorithm. Note that this module is heavy on processing time. FORCE L2IMP includes the programs force-parameter-l2imp to generate an empty parameter file, and force-l2imp to perform the data fusion.

- FORCE WVDB

    - Collection 6 data were removed from LAADS' servers, leaving force-lut-modis nonfunctional. The code was updated to make use of collection 6.1.
    - Due to new NASA policies, LAADS' FTP was shut down, leaving force-lut-modis nonfunctional. The code was updated to make use of the HTTPS server.

- The water vapor climatology has a new column: standard deviation of monthly water vapor. This information is just for estimating the variability for each coordinate/month, and so for the usability of the climatology. This information is not used by FORCE L2PS, thus the old tables can still be used.
- We compiled a ready-to-use, global water vapor database. The dataset is comprised of daily global water vapor data for February 2000 to July 2018 for each land-intersecting Worldwide Reference System 2 (WRS-2) scene, as well as a monthly climatology that can be used if no daily value is available. The dataset is freely available at https://doi.pangaea.de/10.1594/PANGAEA.893109 under the terms of the CC BY 3.0 license. This dataset may relieve you of the burden to generate the water vapor database on your own.
- Fixed a bug in force-lut-modis in finding intersecting MODIS granules.

- FORCE AUX

    - Added new program 'force-mosaic', which generates virtual mosaics from tiled products in Virtual Dataset format. This tool can be used with any tiled FORCE output, i.e. Level 2 / Level 3 / TSA / CSO / ImproPhe. force-mosaic mosaicks everything with the same basename. It Will likely fail for products that have different number of bands in different tiles; this especially applies to the TSS product of force-tsa. The previous, per-module mosaic tools are deprecated.
    - There is a new mandatory parameter in force-qai-inflate, which specifies the output format, i.e. ENVI or GTiff.
    - The usage of force-tile-finder has slightly changed. Before, the Level 2 parameter file needed to be given as input (containing the definition of projection and grid). Now, each data cube is accompanied by a spatial data cube definition, Therefore, only the directory of any gridded data cube needs to be given (containing a data cube definition file). An additional parameter ‘resolution’ was added to relate geographic coordinates to pixel positions, denoting that products with multiple resolutions may be present in a data cube.
    - The usage of force-tabulate-grid has slightly changed. Before, the Level 2 parameter file needed to be given as input (containing the definition of projection and grid). Now, each data cube is accompanied by a spatial data cube definition, Therefore, only the directory of any gridded data cube needs to be given (containing a data cube definition file).

VI. SOFTWARE COMPONENTS

FORCE is a software framework and consists of several executables, which are organized in software components. Table 1 summarizes all modules and programs with a short description. A more detailed description can be found in the corresponding subsections. A brief usage message is displayed when executing a program without arguments. A number of auxiliary files are needed (or are optional) and must be prepared as specified in section VII. It is advised to give the full paths to directories and files wherever applicable.

Table 1. FORCE components.

Component	Program	Short description	Page
FORCE L1AS	The FORCE Level 1 Archiving Suite is intended to assist in organizing and maintaining a clean and consistent Level 1 data pool, downloading of Sentinel-2 data, and maintaining file queues.		18
	force-level1-landsat	Maintenance of Level-1 data pool, Landsat	
	force-level1-sentinel2	Download of data and maintenance of Level-1 data pool, Sentinel-2	
	force-level1-sentinel2-long	Download of data (deprecated long naming convention) and basic maintenance of Level-1 data pool, Sentinel-2	
FORCE L2PS	The FORCE Level 2 Processing System is intended to generate Analysis Ready Data (ARD), i.e. harmonized, standardized and radiometrically consistent Level 2 products. This includes cloud and cloud shadow detection, radiometric correction and data cubing.		21
	force-parameter-level2	Generation of parameter file skeleton for Level 2 processing	
	force-level2	Level 2 processing of image archives	
	force-l2ps	Level 2 processing of single image	
	force-quicklook-level2	Generation of quicklooks of Level 2 products	
FORCE L3PS	The FORCE Level 3 Processing System is intended to generate temporal aggregations of Level 2 data, i.e. pixel-based composites as well as temporal statistics.		31
	force-parameter-level3	Generation of parameter file skeleton for Level 3 processing	
	force-level3	Level 3 processing	
	force-quicklook-level3	Generation of quicklooks of Level 3 products	
FORCE TSA	The FORCE Time Series Analysis component is intended to provide out-of-the-box time series preparation and analysis functionality.		35
	force-parameter-tsa	Generation of parameter file skeleton for time series processing	
	force-tsa	Time series generation / analysis	
FORCE CSO	FORCE Clear Sky Observations (FORCE CSO) is intended for data availability mining.		44
	force-parameter-cso	Generation of parameter file skeleton for CSO processing	
	force-cso	Clear Sky Observations mining	

FORCE ImproPhe	The Force Improving Phenology component is intended to improve the spatial resolution of coarse continuous fields like Land Surface Phenology to Level 2 ARD resolution.		48
	force-parameter-improphe	Generation of parameter file skeleton for ImproPhe processing	
	force-improphe	Improve spatial resolution of coarse continuous fields	
FORCE L2IMP	The Force Level 2 ImproPhe component is intended to increase the spatial resolution of lower resolution Level 2 ARD using higher resolution Level 2 ARD, e.g. to improve the spatial resolution of 30m Landsat imagery to 10m using Sentinel-2 data as targets.		51
	force-parameter-l2imp	Generation of parameter file skeleton for Level 2 ImroPhe processing	
	force-l2imp	Improve spatial resolution of lower resolution Level 2 ARD	
FORCE WVDB	The FORCE Water Vapor Database component can be used to generate and maintain a water vapor database used for atmospheric correction of Landsat data.		54
	force-lut-modis	Generation and maintenance of water vapor database using MODIS products	
FORCE AUX	The FORCE Auxiliary Functionality component is intended to provide small helper programs for specific purposes.		56
	force	Print version and short disclaimer	
	force-mosaic	Virtual mosaicking of data products	
	force-qai-inflate	Inflate QAI bit layers	
	force-tile-finder	Convert geographic coordinates to tile and pixel coordinates	
	force-tabulate-grid	Extract the processing grid as ESRI shapefile	

A. FORCE L1AS - Level 1 Archiving Suite

The FORCE Level 1 Archiving Suite (FORCE L1AS) is intended to assist in organizing and maintaining a clean and consistent Level 1 data pool, as well as downloading of Sentinel-2 data. It is attempted to reduce redundancy, e.g. by removing old data if new processing versions are available. In addition, FORCE L1AS assists in building and updating the file queues needed for the Level 2 processing (see section VII.B).

A glimpse of what you get:

```
2018-09-19 09:06:30 - Found 100 S2A/B files.
/data/Dagobah/edc/l 100%[===================>]  78.46M  18.7MB/s    in 4.4s
/data/Dagobah/edc/l 100%[===================>]  96.50M  19.4MB/s    in 4.4s
/data/Dagobah/edc/l 100%[===================>] 795.83M  33.3MB/s    in 21s
/data/Dagobah/edc/l 100%[===================>] 803.64M  43.4MB/s    in 22s
/data/Dagobah/edc/l 100%[===================>]  78.45M  34.6MB/s    in 2.3s
/data/Dagobah/edc/l 100%[===================>] 734.62M  44.0MB/s    in 17s
/data/Dagobah/edc/l 100%[===================>] 797.96M  44.4MB/s    in 18s
/data/Dagobah/edc/l 100%[===================>] 779.89M  35.2MB/s    in 22s
/data/Dagobah/edc/l 100%[===================>] 814.50M  45.3MB/s    in 19s
/data/Dagobah/edc/l 100%[===================>] 798.21M  39.8MB/s    in 19s
/data/Dagobah/edc/l 100%[===================>]  78.63M  35.2MB/s    in 2.2s
/data/Dagobah/edc/l 100%[===================>]  64.21M  34.7MB/s    in 1.9s
/data/Dagobah/edc/l 100%[===================>] 700.51M  31.1MB/s    in 24s
/data/Dagobah/edc/l 100%[===================>] 292.61M  46.0MB/s    in 7.2s
/data/Dagobah/edc/l 100%[===================>]  79.85M  37.2MB/s    in 2.1s
/data/Dagobah/edc/l 100%[===================>] 749.75M  46.0MB/s    in 18s
/data/Dagobah/edc/l 100%[===================>] 786.91M  41.0MB/s    in 19s
/data/Dagobah/edc/l 100%[===================>] 759.81M  26.2MB/s    in 21s
/data/Dagobah/edc/l 100%[===================>] 540.22M  42.0MB/s    in 13s
/data/Dagobah/edc/l 100%[===================>] 763.79M  44.0MB/s    in 17s
/data/Dagobah/edc/l 100%[===================>] 788.64M   989 B/s    in 31s
/data/Dagobah/edc/l 100%[===================>] 787.45M  29.2MB/s    in 38s
/data/Dagobah/edc/l 100%[===================>] 751.16M  6.75KB/s    in 53s
VS/S2B_MSIL1C_20180  82%[===============>    ] 363.39M  37.2MB/s    eta 3s
```

Fig. 1. Screenshot of downloading Sentinel-2 A/B data.
This is admittedly the most boring FORCE module ;-)

1) Landsat

FORCE can process Level 1 Landsat data, generated using the Level 1 Product Generation System (LPGS) of the U.S. Geological Survey (USGS). Data processed with the outdated National Land Archive Production System (NLAPS) are not supported. At the time of writing, pre-collection and collection 1 data were successfully digested by FORCE. Following Landsat sensors are supported:

- Landsat 4 Thematic Mapper
- Landsat 5 Thematic Mapper
- Landsat 7 Enhanced Thematic Mapper +
- Landsat 8 Operational Land Imager

Before getting started, the full resolution Landsat images must be acquired from the USGS archive, e.g. through EarthExplorer or GloVis; FORCE currently does not provide functionality to download Landsat data. It is recommended to store the compressed images in a consistent data pool (without duplicates or different processing versions). A file queue needs to be prepared (see section VII.B). Both tasks can be handled by *force-level1-landsat*. Extraction of the *.tar.gz archives is not necessary at this point as this is done on the fly during Level 2 processing.

Module	|	force-level1-landsat
Usage	|	force-level1-landsat from Level-1-datapool queue cp/mv [dry]

The input directory (1st argument) is recursively scanned for *.tar.gz files and the Path/Row is extracted from the file path. White-space characters (e.g. when using the Bulk Download Application from USGS) are not allowed in the file path. A subdirectory for every Path/Row is created in the Level 1 directory (2nd argument). The Level 1 archives are moved or copied (4th argument) to the target directory. Make sure to have writing permission in both directories. Files are not imported if they are duplicates of already existing files. Existing files are replaced by files with a newer production number. The 5th argument is optional, and if 'dry' is specified, no files are moved/copied; *force-level1-landsat* only prints what it would do.

A file queue (e.g. named 'level1-landsat-germany.pool') needs to be given as 3rd argument. If it does not exist, it will be created. If it exists, new imports are appended to this file. Outdated files (older production number) are removed from this queue, and the new imports are appended to the end. This queue is needed for Level 2 processing; all images with 'QUEUED' status will be processed, then set to 'DONE'.

Note that the input directory should not be a parent of the target directory. The input directory is scanned recursively, thus files already in the L1 data pool will be moved again if the target is a child of the input (this is unnecessary and might take a while).

2) Sentinel-2

FORCE can process Level 1C Sentinel-2 A/B data as provided by ESA through their data hub. At the time of writing, processing baseline 02.06 data was successfully digested by FORCE.

Before getting started, the full resolution Sentinel-2 images must be acquired from the ESA archive. It is recommended to store the unzipped images in a consistent data pool (without duplicates or different processing versions). A file queue needs to be prepared (see section VII.B). Downloading data, unzipping, managing the data pool, and preparing/updating the file queue can be handled by *force-level1-sentinel2* – and with slightly reduced functionality – by the deprecated *force-level1-sentinel2-long*.

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

The Level 1 images are downloaded to the Level 1 data pool (1st argument), which should be given as absolute and full file path. Files are not downloaded/imported if they are duplicates of already existing files. Existing files are replaced by files with a newer production number.

A file queue (e.g. named ‘level1-sentinel2-germany.pool’) needs to be given as 2nd argument. If it does not exist, it will be created. If it exists, new imports are appended to this file. Outdated files (older production number) are removed from this queue, and the new imports are appended to the end. This queue is needed for Level 2 processing; all images with ‘QUEUED’ status will be processed, then set to ‘DONE’.

Each acquisition covered by the bounding box (3rd argument) is downloaded/imported. The bounding box (square or not) encloses your study area and must be given as “X1/Y1,X2/Y2,X3/Y3,...,X1/Y1”. The box must be closed (first X/Y = last X/Y). X/Y must be given in decimal degree with negative values for West and South coordinates.

Starttime and endtime (4th and 5th argument) specify a temporal subset and refer to the acquisition time. The dates must be given in YYYY-MM-DD.

Min-cc and max-cc (6th and 7th argument) are the minimum and maximum allowed cloud percentage as integer values ranging from 0 to 100.

At the end of 2016, ESA changed the data format of their Level 1C Sentinel-2 processing (shorter file names and only one tile per product). The program *force-level1-sentinel2-long* is the deprecated version of *force-level1-sentinel2*, and is intended for obtaining older data with the long-naming convention (which is still distributed by ESA). For new acquisitions (i.e. short naming convention), *force-level1-sentinel2* should be used (*force-level1-sentinel2-long* will be removed from the repository once ESA will have reprocessed all data – at time of writing, reprocessing is in progress – but moving slowly). Due to the more complex data structure of the long naming convention, the version checking is non-functional in *force-level1-sentinel2-long*, thus duplicates might be present.

Data will be downloaded from ESA’s API Hub, which requires an account at ESA’s end. On your end, your login credentials must be placed in a hidden file ‘.scihub’ in your home directory (you can chmod to 400), with user name in the first line and password in the second line (special characters might be problematic).

It is possible to call *force-level1-sentinel2* (and *force-level2*) from a cronjob, in which case a near-real time processing can be realized.

B. FORCE L2PS - Level 2 Processing System

The FORCE Level 2 Processing System (FORCE L2PS) is intended to generate harmonized, standardized and radiometrically consistent Level 2 products (→ ARD: Analysis Ready Data) based on Level 1 data as provided from the space agencies. This includes cloud and cloud shadow detection, radiometric correction and data cubing. This section summarizes the usage of FORCE L2PS, its helper programs and the output format.

A glimpse of what you get:

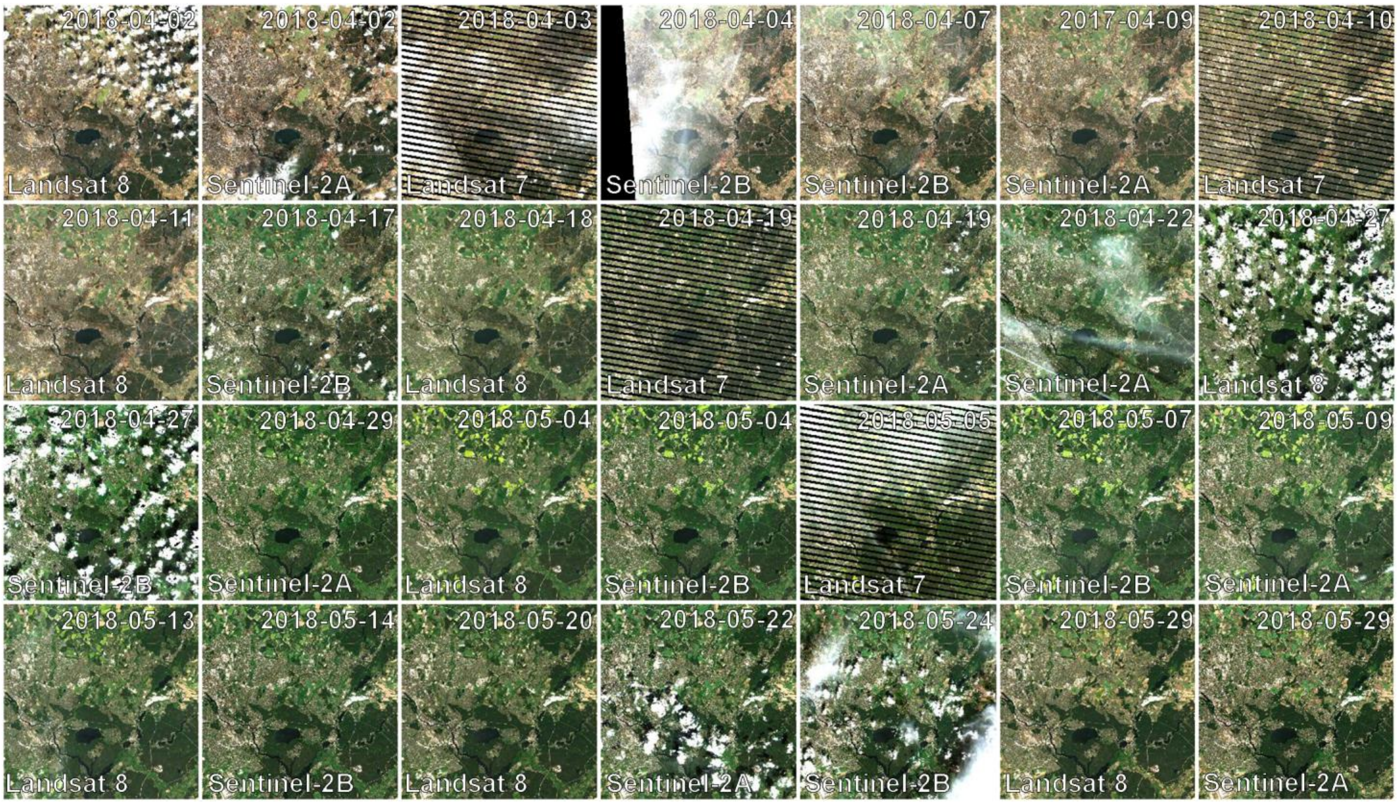

Fig. 2. Data Cube of Landsat 7/8 and Sentinel-2 A/B ARD.
A two-month period of atmospherically corrected imagery acquired over South-East Berlin, Germany, is shown here.
[Images were corrected using force-level2, true-color quicklooks were generated using force-quicklook-level2]

1) Parameterization

A parameter file is mandatory for FORCE L2PS; *force-parameter-level2* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (see usage). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.A summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-level2
Usage	|	force-parameter-level2 dir

There is a small bug in *force-parameter-level2* in FORCE v. 2.0: one line is missing in the generated parameter file. The projection must be given in two lines (see section VII.A), and the line containing the PROJECTION tag is missing. Place following line before the projection definition.

PROJECTION =

2) Digital Elevation Model

A digital elevation model (DEM) is recommended for FORCE L2PS (it is used for cloud detection, topographic and atmospheric correction). The user should provide a DEM that covers the complete study area at adequate spatial resolution. The DEM must be specified in the parameter file (see section VII.A). There is not really a restriction on the source, projection or resolution; the file format must be supported by *GDAL*. The DEM will be warped to the extent and resolution of the processed image using bilinear resampling. The DEM needs to be provided in meters above sea level with nodata -32767.

The user can choose to process without a DEM; in this case the surface is assumed flat @ z = 0m. Topographic correction cannot be used without a DEM. The quality of atmospheric correction and cloud /cloud shadow detection will suffer without DEM.

3) Processing

(a) Image archives

The core module of FORCE L2PS is *force-level2*, which handles the mass processing of the Level 1 input. In principle, the same algorithm is applied to all sensors – although specific processing options are triggered or are only available for some sensors (e.g. Sentinel-2 resolution merge). Each image enqueued in the file queue (see section VII.B) is processed to Level 2 according to the specifications in the parameter file (see section VII.A); the file queue is specified in the parameter file too (the filename of the queue, not the content of the file). After processing, the image is dequeued, and as such it is possible to schedule processing in near-real time, e.g. if called from a cronjob. If reprocessing is required, the image needs to be enqueued again (see section VII.B). The processed images, as well as a logfile are written to the output directory given in the parameter file.

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

The parameter file needs to be given as 1st argument. The number of CPUs used for parallel processing needs to be given as 2nd argument. Each processor handles one input image. Note that *GNU parallel* spawns jobs by using multiple ssh logins; the number of allowed ssh logins per user must be sufficiently large (ask your system administrator to increase this number if necessary). It is advised to monitor the workload. Swapping should be prevented – in this case, decrease the number of parallel jobs. The number of CPUs can be re-adjusted while(!) force-level2 is running. A file 'cpu-$TIME' is temporarily created in DIR_TEMP. This file can be modified. Note that the effect is not immediate, as the load is only adjusted after one of the running jobs (images) is finished.

To prevent an I/O jam at startup (by reading / extracting a lot of data simultaneously), a delay (in seconds) needs to given as 3rd argument. Each 'delay' seconds, the processing of a new image is started until 'ncpu' parallel jobs are running. Depending on processing speed per image plus 'ncpu' and 'delay' settings, it is possible that 'ncpu' parallel processes won't be reached.

The necessary delay is dependent on your system's architecture (I/O speed etc), and on sensor to be processed. Our recommendations are as follows (but note that these values might not be fitting for you, each system is different):

- Sentinel-2: large delay (large data volume + fairly long processing time). Try 20 seconds.

- Landsat: delay should approximately reflect the time necessary for extracting a single *.tar.gz archive. Small values for Landsat 5-7 are reasonable (small data volume + short processing time). Try 3–5 seconds. Larger delays may be necessary for Landsat 8 (fairly high data volume + short processing time). Try 10–15 seconds.

(b) Single images

The workhorse of FORCE L2PS is *force-l2ps*, which is a lower-level routine called from within *force-level2*. It processes one single image. For the majority of users, it is recommended to use *force-level2* instead of directly calling *force-l2ps*. However, for specific purposes (e.g. testing / debugging), the expert user may want to use this program directly (or if you want to implement your own job scheduler).

Module	|	force-l2ps
Usage	|	force-l2ps L1-image par-file

The 1st argument is the directory that contains the image data. In case of Landsat, the *.tar.gz archive needs to be extracted before processing. In case of Sentinel-2, the *.zip archive needs to extracted before processing and one tile (directory) within the 'GRANULE' directory must be given as input. Note that the extraction of Landsat images is automatically performed within *force-level2*, and the extraction of Sentinel-2 data is performed within *force-level1-sentinel2* (see VI.A.2). The parameter file needs to be given as 2nd argument.

The direct usage of *force-l2ps* is recommended for debugging or for detailed output. The debugging mode also features extensive output where images for most processing steps are saved. Note that these images are intended for software development and do not necessarily have intuitive file names; metadata or projections are also not appended. If debug output is required, the software needs to be re-compiled, which should only be done by expert users. If DEBUG is activated, *force-level2* does not allow you to process multiple images or to use parallel processing (your system will be unresponsive because too much data is simultaneously written to the disc, and parallel calls to *force-l2ps* would overwrite the debugging images). For debugging, follow these steps:

- Modify force/src/force.h
- After the includes, there is a 'compiler options' section. Remove the comments before DEBUG and DEBUGPATH and define an existing path for DEBUGPATH. The debug images will be saved in this directory.
- Re-compile the software

4) Quicklooks

Quicklooks of the processed Level 2 data can be generated using *force-quicklook-level2*. It searches files that match the pattern '*BOA*.dat/tif' in the directory given as 1st argument (typically the output directory of FORCE L2PS), and creates a quicklook in the directory given as 2nd argument. The data organization of the Level 2 archive will be mirrored in the quicklook archive (tiles or original spatial reference system). Existing quicklooks are skipped (to enable efficient and scheduled processing, e.g. if called from a cronjob). The true color quicklooks have a fixed stretch from 0–maxval (3rd argument). The upper limit of the stretch needs to be given in scaled reflectance (scaling factor 10,000), e.g. 1500. This value can be decreased/increased for very dark/bright landscapes. The quicklooks are generated with *gdal_translate* in JPEG format and have a fixed size of 256 x 256 px.

Module	|	force-quicklook-level2
Usage	|	force-quicklook-level2 L2-archive QUICKLOOK-archive maxval

5) Output format

(a) Data organization

Depending on parameterization, the output data are organized according to their original spatial reference system (WRS-2 frames / MGRS zones) or are provided in a gridded data structure as ARD (strongly recommended!). The tiles (or original reference system) manifest as directories in the file system, and the images are stored within. The user can choose to keep the original projection (UTM) or to reproject all data to one consistent projection (strictly recommended for ARD!). See section VII.H for more details.

(b) Naming convention

Following 29-digit naming convention is applied to all output files:

```
20160823_LEVEL2_SEN2A_BOA.tif
20160823_LEVEL2_SEN2A_BOA.hdr
```

Digits 1–8 Acquisition date as YYYYMMDD

Digits 10–15 Product Level

Digits 17–21 Sensor ID

LND04	Landsat 4 Thematic Mapper
LND05	Landsat 5 Thematic Mapper
LND07	Landsat 7 Enhanced Thematic Mapper +
LND08	Landsat 8 Operational Land Imager
SEN2A	Sentinel-2A MultiSpectral Instrument
SEN2B	Sentinel-2B MultiSpectral Instrument

Digits 23–25 Product Type

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)

Digits 27–29 File extension

tif	image data in compressed GeoTiff format
dat	image data in flat binary ENVI format
hdr	metadata

(c) Product type

Reflectance data (BOA / TOA) and Quality Assurance Information (QAI) are standard output and cannot be disabled.

AOD / CLD / WVP / VZN / HOT output are optional, images are single-band at the same resolution as BOA / TOA.

⇒ *BOA / TOA reflectance*

Bottom-of-Atmosphere (BOA) reflectance is standard output if atmospheric correction is used. Top-of-Atmosphere (TOA) reflectance is standard output if atmospheric correction is not used. BOA / TOA data contain multiple bands (≙ wavelengths, see metadata and following tables). All bands are provided at the same spatial resolution (set by RESOLUTION parameter). Bands intended for atmospheric characterization are not output (e.g. ultra-blue, water vapor or cirrus bands). Following tables summarize the output bands for each sensor.

Table 2. Landsat 4–5 Thematic Mapper (TM) bands.
Summary of USGS band definitions, and corresponding output bands of FORCE L2PS.

USGS Level 1 band name	Wavelength name	Wavelength range in µ	Resolution in m	FORCE Level 2 output band
1	Blue	0.45–0.52	30	1
2	Green	0.52–0.60	30	2
3	Red	0.63–0.69	30	3
4	Near Infrared	0.76–0.90	30	4
5	Shortwave Infrared 1	1.55–1.75	30	5
6	Thermal Infrared	10.40–12.50	30 (120*)	-**
7	Shortwave Infrared 2	2.08–2.35	30	6

* Band is acquired at 120m resolution, but USGS products are resampled and provided at 30m.
** Thermal band is used internally for cloud / cloud shadow detection, but not output.

Table 3. Landsat 7 Enhanced Thematic Mapper Plus (ETM+) bands.
Summary of USGS band definitions, and corresponding output bands of FORCE L2PS.

USGS Level 1 band name	Wavelength name	Wavelength range in µ	Resolution in m	FORCE Level 2 output band
1	Blue	0.45–0.52	30	1
2	Green	0.52–0.60	30	2
3	Red	0.63–0.69	30	3
4	Near Infrared	0.77–0.90	30	4
5	Shortwave Infrared 1	1.55–1.75	30	5
6	Thermal Infrared	10.40–12.50	30 (60*)	-**
7	Shortwave Infrared 2	2.09–2.35	30	6
8	Panchromatic	0.52–0.90	15	-

* Band is acquired at 60m resolution, but USGS products are resampled and provided at 30m.
** Thermal band is used internally for cloud / cloud shadow detection, but not output.

Table 4. Landsat 8 Operational Land Imager (OLI) / Thermal Infrared Sensor (TIRS) bands. Summary of USGS band definitions, and corresponding output bands of FORCE L2PS.

USGS Level 1 band name	Wavelength name	Wavelength range in µ	Resolution in m	FORCE Level 2 output band
1	Ultra-Blue	0.435–0.451	30	-**
2	Blue	0.452–0.512	30	1
3	Green	0.533–0.590	30	2
4	Red	0.636–0.673	30	3
5	Near Infrared	0.851–0.879	30	4
6	Shortwave Infrared 1	1.566–1.651	30	5
7	Shortwave Infrared 2	2.107–2.294	30	6
8	Panchromatic	0.503–0.676	15	-
9	Cirrus	1.363–1.384	30	-***
10	Thermal Infrared 1	10.60–11.19	30 (100*)	-****
11	Thermal Infrared 2	11.50–12.51	30 (100*)	-

* Bands are acquired at 100m resolution, but USGS products are resampled and provided at 30m.
** Ultra-Blue band is used internally for aerosol retrieval, but not output.
*** Cirrus band is used internally for cirrus cloud detection, but not output.
**** Thermal band is used internally for cloud / cloud shadow detection, but not output.

Table 5. Sentinel-2 A/B MultiSpectral Instrument (MSI) bands. Summary of ESA band definitions, and corresponding output bands of FORCE L2PS.

ESA Level 1 band name	Wavelength name	Wavelength range in µ	Resolution in m	FORCE Level 2 output band
1	Ultra-Blue	0.430–0.457	60	-*
2	Blue	0.440–0.538	10	1
3	Green	0.537–0.582	10	2
4	Red	0.646–0.684	10	3
5	Red Edge 1	0.694–0.713	20	4
6	Red Edge 2	0.731–0.749	20	5
7	Red Edge 3	0.769–0.797	20	6
8	Broad Near Infrared	0.760–0.908	10	7
8A	Near Infrared	0.848–0.881	20	8
9	Water Vapor	0.932–0.958	60	-**
10	Cirrus	1.337–1.412	60	-***
11	Shortwave Infrared 1	1.539–1.682	20	9
12	Shortwave Infrared 2	2.078–2.320	20	10

* Ultra-Blue band is used internally for aerosol retrieval, but not output.
** Water vapor band is used internally for water vapor retrieval, but not output.
*** Cirrus band is used internally for cirrus cloud detection, but not output.

⇒ *Quality Assurance Information*

Quality Assurance Information (QAI product) are key for any higher-level analysis of ARD. This product contain the cloud masks etc. **USE QAI RIGOROUSLY!!!** QAI are provided bit-wise for each pixel, thus the integers have to be parsed using following convention (see also *force-qai-inflate* in section VI.I.2). As an example, integer 28672 would be a poorly illuminated, sloped pixel where water vapor could not have been estimated.

Bit:	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	
Flag:	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	∑ = 28672

Nodata values are values where no data was observed, where no auxiliary information was given (nodata values in DEM), or where data is substantially corrupt. The latter case includes impulse noise (for Landsat 5–7), or pixels that would end up with reflectance > 2.0 or reflectance < -1.0.

Clouds are automatically detected, 100% accuracy for any given image cannot be given. In general, we tried to balance cloud masking, but we believe that it is most important to reduce omission errors for automated processing and analysis of large image archives. Therefore, commission error is probably higher than omission error for any given scene (over time, we expect this to level out on the pixel scale). Opaque clouds (confident cloud) are buffered by 300m (less confident cloud). Cirri are only detected for Landsat 8 and Sentinel-2, using thresholds of the cirrus and blue band.

Cloud shadows are detected on the basis of the cloud layer. If a cloud is missed, the cloud shadow is missed, too. If a false positive cloud is detected, false positive cloud shadows follow. As cloud shadow detection is of less accuracy, all high shadow matches are included in the cloud shadow mask, therefore commission error is larger than in the cloud mask.

Cloud, cloud shadow, snow and water flags are exclusive. A pixel cannot have multiple flags. Water takes precedence over snow. Snow takes precedence over cloud. Note that snow and cold clouds can be easily confused. No shadow is matched for snow pixels. Opaque clouds, and buffered clouds take precedence over cirrus clouds. Clouds take precedence over cloud shadows.

It is advised to always filter for snow, clouds, and cloud shadows (unless you are specifically interested in one of them).

Aerosol Optical Depth is estimated for fairly coarse grid cells. If there is no valid AOD estimation in any cell, values are interpolated. If there is no valid AOD estimation for the complete image, a fill value is assigned (AOD is guessed). If AOD @550nm is higher than 0.6, it is flagged as high aerosol; this is not necessarily critical, but should be used with caution (see subzero flag).

If reflectance in any band is < 0, the subzero flag is set. This can point to overestimation of AOD. Depending on application, you should use this data with caution.

If DNs were saturated, or if reflectance in any band is > 1, the saturation flag is set. Depending on application, you should use this data with caution.

If sun elevation is smaller than 15°, the high sun zenith flag is set. Use this data with caution, radiative transfer computations might be out of specification.

The illumination state is related to the quality of the topographic correction. If the incidence angle is smaller than 55°, quality is best. If the incidence angle is larger than 80°, the quality of the topographic correction is low, and data artefacts are possible. If the area is not illuminated at all, no topographic correction is done (values are the same as without topographic correction).

The slope flag indicates whether a simple cosine correction (slope ≤ 2°) was used for topographic correction, or if the enhanced C-correction was used (slope > 2°).

The water vapor flag indicates whether water vapor was estimated, or if the scene average was used to fill. Water vapor is not estimated over water and cloud shadow pixels. This flag only applies to Sentinel-2 images.

Table 6. Quality Assurance Information (QAI) description.
[continued on next pages]

Bit No.	Parameter Name	Bit comb.	Integer	State
0	Valid data	0	0	valid
		1	1	no data

1-2	Cloud state	00	0	clear
		01	1	less confident cloud (i.e. buffered cloud)
		10	2	confident, opaque cloud
		11	3	cirrus
3	Cloud shadow flag	0	0	no
		1	1	yes
4	Snow flag	0	0	no
		1	1	yes
5	Water flag	0	0	no
		1	1	yes
6-7	Aerosol state	00	0	Estimated (best quality)
		01	1	interpolated (mid quality)
		10	2	high (might or might not be problematic, watch out)
		11	3	fill (use with caution, AOD estimate is just a guess)
8	Subzero flag	0	0	no
		1	1	yes (use with caution)
9	Saturation flag	0	0	no
		1	1	yes (use with caution)
10	High sun zenith flag	0	0	no
		1	1	yes (use with caution)
11-12	Illumination state	00	0	good (best quality for topographic correction)
		01	1	low (good quality for topographic correction)
		10	2	poor (low quality for topographic correction, artefacts are possible)
		11	3	shadow (no topographic correction applied)

13	Slope flag	0	0	no (< 2° slope, topogr. correction: cosine correction applied)
		1	1	yes (> 2° slope, topogr. correction: enhanced C-correction applied)
14	Water vapor flag	0	0	measured (best quality, only used for Sentinel-2)
		1	1	fill (scene average, e.g. over water, only used for Sentinel-2)
15	Empty	0	0	TBD

⇒ *Aerosol Optical Depth*

The Aerosol Optical Depth (AOD) product is optional output. It contains the AOD of the green band (~550 nm). This product is not used by any of the higher-level FORCE modules.

⇒ *Cloud / cloud shadow / snow distance*

The Cloud / cloud shadow / snow distance (CLD) product is optional output. The cloud distance gives the distance to the next opaque cloud, buffered cloud, cirrus cloud, cloud shadow or snow. This product can be used in FORCE L3PS to generate Best Available Pixel (BAP) composites.

Note that this is not the actual cloud mask! For cloud masks and quality screening, rather use the QAI product.

⇒ *Water vapor*

The Water vapor (WVP) product is optional output. It contains the atmospheric water vapor (as derived from the Sentinel-2 data on pixel level, or as ingested with the water vapor database for Landsat). This product is not used by any of the higher-level FORCE modules.

⇒ *View zenith*

The View zenith (VZN) product is optional output. It contains the view zenith (the average view zenith for Sentinel-2, and an approximated view zenith for Landsat). This product can be used in FORCE L3PS to generate Best Available Pixel (BAP) composites.

⇒ *Haze Optimized Transformation*

The Haze Optimized Transformation (HOT) product is optional output. It contains the HOT index, which is computed on TOA reflectance (and therefore cannot be computed on Level 2 ARD). The HOT is useful to avoid hazy and residual cloud contamination. This product can be used in FORCE L3PS to generate Best Available Pixel (BAP) composites.

(d) File format

The data are provided in (i) ENVI Standard format (flat binary images), or (ii) as GeoTiff (LZW compression with horizontal differencing). Each dataset consists of an image dataset (.dat/,tif) and additional metadata (.hdr). The image data have signed 16bit datatype and band sequential (BSQ) interleaving.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info etc. Additional information like acquisition time (including date and time), cloud cover per image and tile, sun position and processing information are also provided.

(e) *Logfile*

A logfile is created by *force-level2* in the output directory. Following 29-digit naming convention is applied:

```
FORCE-L2PS_20170712040001.log
```

Digits 1–10 Processing module

Digits 12–25 Processing time (start time) as YYYYMMDDHHMMSS

Digits 27–29 File extension

Typical entries look like this:

```
LC08_L1TP_195023_20180110_20180119_01_T1: sc:   0.10%. cc:  89.59%. AOD: 0.2863.
# of targets: 0/327.  4 product(s) written. Success! Processing time: 32 mins 37
secs
LC08_L1TP_195023_20170328_20170414_01_T1: sc:   0.00%. cc:   2.56%. AOD: 0.0984.
# of targets: 394/6097.  6 product(s) written. Success! Processing time: 19 mins
03 secs
LC08_L1TP_195023_20170312_20170317_01_T1: sc:   0.29%. cc:  91.85%. Skip.
Processing time: 13 mins 22 secs
```

The first entry indicates the image ID, followed by overall snow and cloud cover, aerosol optical depth @ 550 nm (scene average), the number of dark targets for retrieving aerosol optical depth (over water/vegetation), the number of products written (number of tiles, this is dependent on tile cloud cover, and FILE_TILE), and a supportive success indication. In the case the overall cloud coverage is higher than allowed, the image is skipped. The processing time (real time) is appended at the end.

C. FORCE L3PS - Level 3 Processing System

The FORCE Level 3 Processing System (FORCE L3PS) is intended to generate temporal aggregations of Level 2 data to provide seamless, gap free, and highly Analysis Ready Data (hARD) over very large areas. This includes pixel-based composites as well as spectral temporal statistics (e.g. average reflectance within compositing period). hARD are the optimal input for many machine learning algorithms, e.g. for land cover /change classification purposes. Pixel-based composites can either be static or phenology-adaptive. In the latter case, a Land Surface Phenology (LSP) dataset is required (see subsection). This section summarizes the usage of FORCE L3PS, its helper programs and the output format.

FORCE L3PS can only be used with Level 2 ARD.

A glimpse of what you get:

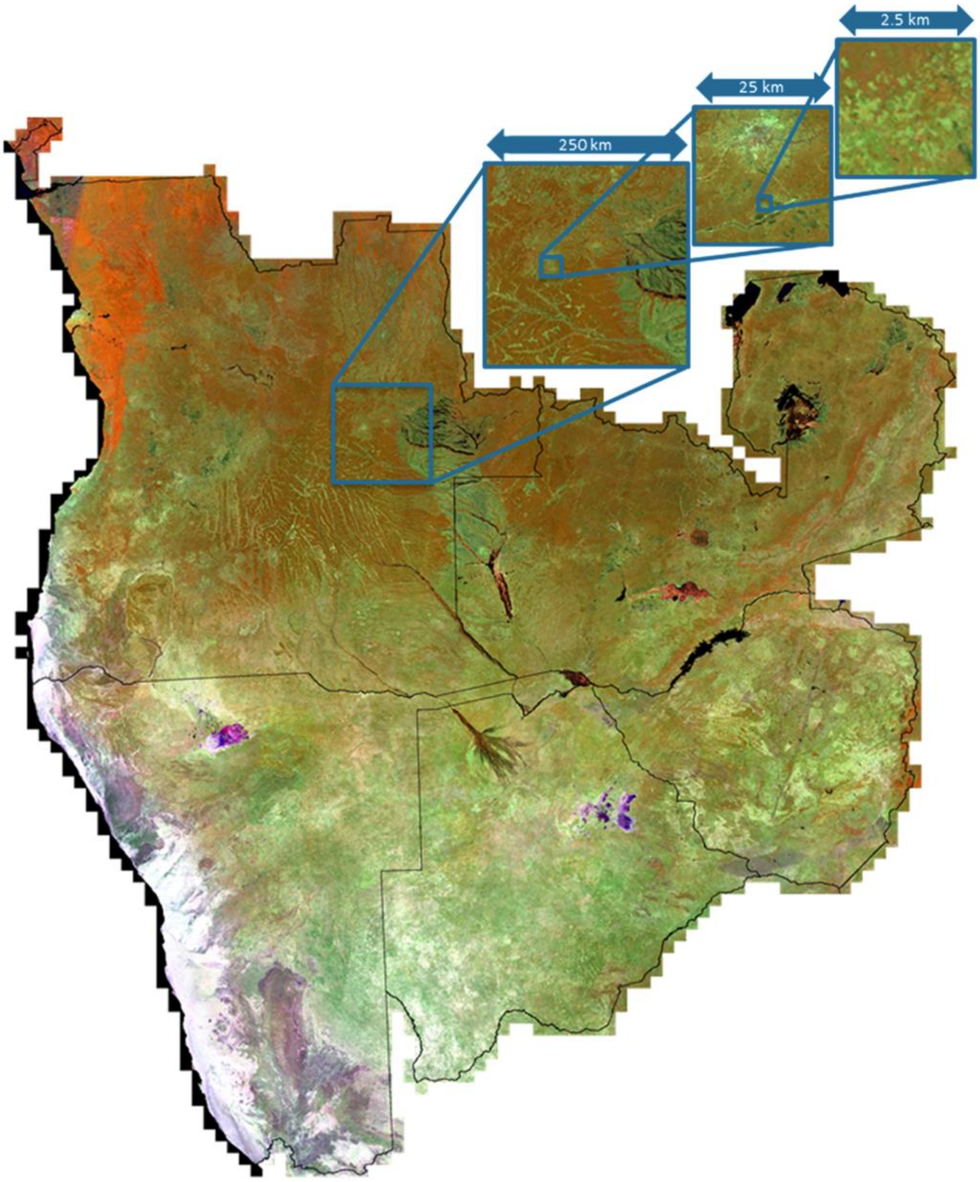

Fig. 3. Phenology Adaptive Composite (PAC) using Landsat 5–7.
The Best Available Pixel (BAP) composite (phenology-adaptive code: End of Season 2018) was computed for Angola, Zambia, Zimbabwe, Botswana and Namibia.
[The composite was generated with force-level3, then mosaicked using force-mosaic]

1) Parameterization

A parameter file is mandatory for FORCE L3PS; *force-parameter-level3* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (1st argument). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.D summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-level3
Usage	|	force-parameter-level3 dir

2) Land Surface Phenology

A Land Surface Phenology (LSP) dataset may be input to generate phenology-adaptive composites – or s.th. similar that is dependent on spatial variation of the target dates. This is optional and may be omitted if static composites or temporal statistics are used. The LSP dataset needs to be prepared in the same grid as the Level 2 data (i.e. in a mirrored data structure). Three (or more) images need to be prepared for each tile, i.e. seasonal parameters describing points in time (e.g. the timing of start of season, peak of season, end of season, …). For compositing, a sequence of three images needs to be selected as temporal target, and the filenames must contain a unique ID (e.g. SOS, POS, EOS; see section VII.D. The data are expected to be in ENVI standard or GeoTiff format with file extension *.dat/tif, *.hdr. Each file is a multi-layer image with years as bands (the first year is specified in the Level 3 parameter file; see section VII.D. Note that missing years are not allowed (use a fill band instead). The values should be in days relative to a custom starting point (see section VII.D. Leap years are not taken into account and each year consists of 365 days.

3) Processing

The core module of FORCE L3PS is *force-level3*, which generates composites and/or spectral temporal statistics. The processed images are written to the output directory given in the Level 3 parameter file (see VII.D for all available options).

Module	|	force-level3
Usage	|	force-level3 par-file

The parameter file needs to be given as 1st argument. All options need to be specified in the parameter file, e.g. tiled or mosaicked output, spatial extent, spatial resolution, temporal target / compositing period, phenology-adaptive or static method etc. (see VII.D for all available options).

At the end of processing, histograms about number of observations, sensors, days used for compositing, etc. is printed to the screen.

4) Quicklooks

Quicklooks of the processed Level 3 data can be generated using *force-quicklook-level3*. It searches files that match the pattern '*BAP*.dat/tif/vrt' in the directory given as 1st argument (typically the output directory of FORCE L3PS), and creates a quicklook in the directory given as 2nd argument. The true color quicklooks have a fixed stretch from 0–maxval (3rd argument). The upper limit of the stretch needs to be given in scaled reflectance (scaling factor 10,000), e.g. 1500. This value can be decreased/increased for very dark/bright landscapes. The quicklooks are generated with *gdal_translate* in JPEG format. The 4th argument needs to be given if the Level 3 products are not tiled (→ physical mosaic or virtual mosaic). If the data are tiled, the quicklooks have a fixed size of 256 x 256 px, otherwise it is 20 % of the mosaic size. The data organization of the Level 3 archive will be mirrored in the quicklook archive. Existing quicklooks are skipped (to enable efficient and scheduled processing, e.g. if called from a cronjob).

Module	|	force-quicklook-level3
Usage	|	force-quicklook-level3 L3-archive QUICKLOOK-archive maxval [--no-tile]

5) Output format

(a) Data organization

Depending on parameterization, the output data are organized in the gridded data structure used for Level 2 processing, or mosaics are generated (covering the complete requested area). The tiles (or a directory named ‘mosaic’) manifest as directories in the file system, and the images are stored within.

(b) Naming convention

Following 29-digit naming convention is applied to all output files:

```
20150415_LEVEL3_MULTI_BAP.tif
20150415_LEVEL3_MULTI_BAP.hdr
```

Digits 1–8 Temporal target for compositing as YYYYMMDD

Digits 10–15 Product Level

Digits 17–21 Set of spectral bands

LNDLG	Landsat legacy bands
SEN2L	Sentinel-2 land surface bands
SEN2H	Sentinel-2 high-res bands
R-G-B	Visual bands

Digits 23–25 Product Type

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)

Digits 27–29 File extension

tif	image data in compressed GeoTiff format
dat	image data in flat binary ENVI format
hdr	metadata

(c) Product type

⇒ *Reflectance data*

Reflectance data (BAP and temporal statistics) contain multiple bands (≙ wavelengths). All bands are provided at the same spatial resolution. Single-sensor composites have the same bands as the corresponding Level 2 data. Multi-sensor composites only contain overlapping bands. Exclusive bands are discarded. Note that no spectral adjustment is made.

⇒ *Compositing information*

The compositing information (INF) product contains information about the selected observation in the BAP product. It is a multi-layer image with following bands:

1	QAI	Quality Assurance Information of BAP (bit coding, nodata: 1)
2	# of obs.	Number of cloud-free observations within compositing period
3	D	Acquisition DOY of BAP
4	Y	Acquisition Year of BAP
5	ΔD	Difference between D and Target DOY
6	Sensor	Sensor ID of BAP (in the order, which was given in the parameter file)

⇒ *Compositing score*

The compositing score (SCR) product contains the scores of the BAP (minimum: 0, maximum: 1, scale: 10000). It is a multi-layer image with following bands:

1	S_T	Total score
2	S_D	DOY score (intra-annual score)
3	S_Y	Year score (inter-annual score)
4	S_C	Cloud distance score
5	S_H	Haze score
6	S_R	Correlation score
7	S_V	View angle score

(d) File format

The data are provided in (i) ENVI Standard format (flat binary images), or (ii) as GeoTiff (LZW compression with horizontal differencing). Each dataset consists of an image dataset (.dat/,tif) and additional metadata (.hdr). The image data have signed 16bit datatype and band sequential (BSQ) interleaving.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info, band names etc.

D. FORCE TSA - Time Series Analysis

FORCE Time Series Analysis (FORCE TSA) is intended to provide out-of-the-box time series preparation and analysis functionality. The user can select from a number of spectral indices or unmix spectra using custom endmembers. FORCE TSA is capable of extracting quality-controlled time series with a number of aggregation and interpolation techniques. Annual Land Surface Phenology metrics can be derived, and change and trend analyses can be performed on any of the generated time series. Many outputs of FORCE TSA are considered as highly Analysis Ready Data plus (hARD+), meaning that generated products can be directly used to fuel your research questions without any further processing. This section summarizes the usage of FORCE TSA, its helper programs and the output format.

FORCE TSA can only be used with Level 2 ARD.

A glimpse of what you get:

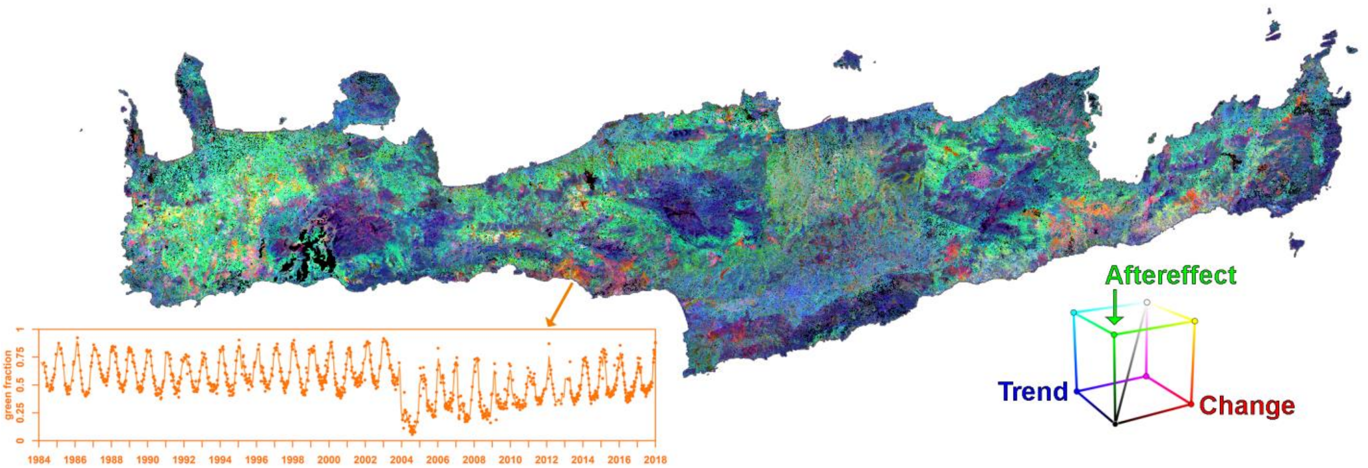

Fig. 4. Phenology-based change and trend analysis.
Change, Aftereffect, Trend transformation (CAT) showing both long-term (30+ years) gradual and abrupt changes over Crete, Greece. The CAT transform was applied to the Value of Base Level (VBL) annual time series, which was itself derived by inferring 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]

1) Parameterization

A parameter file is mandatory for FORCE TSA; *force-parameter-tsa* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (1st argument). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.E summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-tsa	
Usage	|	force-parameter-tsa	dir

2) Processing

The core module of FORCE TSA is *force-tsa*, which extracts time series and performs time series analyses. The processed images are written to the output directory given in the TSA parameter file (see section VII.E for all available options).

Module	|	force-tsa	
Usage	|	force-tsa	par-file

The parameter file needs to be given as 1st argument. All options need to be specified in the parameter file.

3) Quick guide

FORCE TSA is not covered by a scientific publication. Therefore, this section briefly summarizes the key functionality; Fig. 5 depicts the general workflow within FORCE TSA; see also section VII.E for all available options and possible keywords.

The analysis can be constrained to a certain area. The general extent in tile coordinates should be specified, a tile white-list may be used and analysis masks may be input. These need to be prepared as specified in the analysis mask subsection.

The analysis can be performed with observations from all available sensors, or a subset of these. If multiple sensors are used, only overlapping bands are considered. Note that no spectral adjustment is made. The program will exit with an error message if the requested index is not available for the chosen sensor set.

The spatial resolution on which the analysis is performed, needs to be given too. Image decimation/replication is taken care of using nearest neighbor resampling. Alternatively, spatial resolution may be reduced using approximated Point Spread Functions (PSF). The PSFs are parameterized with Full Width at Half Maximum (FWHM) corresponding to the size of one pixel at the target resolution.

A temporal subset can be specified in terms of years, months, and DOYs. The most restrictive seasonal subset is used. As an example, if the min–max days are set to 1–200 and the min–max month are set to 3–12, only acquisitions between DOY 60 and 200 are used. The years can be further subsetted, e.g. a trend analysis can be performed on all July–August images for the years 2000–2010.

In the next step, a spectral band (e.g. NIR), spectral index (e.g. NDVI), or Spectral Mixture Analysis (SMA) is selected. The full (or subset) time series corresponding to this 'index' is generated, quality-controlled, and potentially output. It is also possible to center (or standardize) each pixel's time series to its mean (and standard deviation) before output. All products indicated by a USB-plug in Fig. 5 can be output; all products indicated by * can be centered / standardized before output.

In the case of SMA, a endmember file needs to be provided (see VII.F). Only the abundance image related to one endmember is retained. The SMA can be constrained (Sum-to-One and/or Non-Negativity), and the fractions can be shade normalized. The shade spectrum (photogrammetric zero or measured shade) needs to go into the last column of the endmember file. The RMSE time series of the SMA can be output.

Quality control is in full user control. All provided quality flags (see VI.B.5) can be used individually. Use this option rigorously!

In the next step, a summary of the full time series can be generated. Basic statistics include mean, standard deviation, minimum and maximum.

The time series can be interpolated / smoothed at custom time steps. Currently available are linear interpolation, moving average filter, and Radial Basis Function (RBF) ensembles. The RBF kernel strengths are adapted by weighting with actual data availability within the time period covered by each kernel.

Afterward, the full time series are aggregated before performing a time series analysis. The user can choose to ‘fold’ the full time series by year, month, week or day. If folded by year, the observations are grouped by year to generate an annual time series. If folded by month, the observations are grouped by months, which gives up to 12 values per pixel. If folded by week, the observations are grouped by week numbers (starting with DOY 1), which gives up to 52 values per pixel; note that the last week contains some more days. If folded by day, the observations are grouped by DOY, which gives up to 365 values per pixel. The time series can be folded with mean, minimum or maximum statistics. The user can request to output any of these folded time series, but the following trend analysis can only be performed on one.

Alternatively, the interpolated time series can be folded by year with LSP method, which means that annual Land Surface Phenology (LSP) metrics are extracted. The extraction of LSP is performed by the SPLITS code, which is a spline-based methodology. Note that FORCE needs to be compiled with enabled SPLITS (see IV.F). The number of segments for fitting the splines needs to be specified (in segments per 365 days); more segments give a more detailed fit; the detail vs. smoothing capability of the spline needs to be considered with respect to application, land cover of interest and data density. Deriving LSP will only work nicely if data availability throughout the year is high; do not expect good results for sparse data. To derive LSP metrics for a given year, some data from the previous and next year need to be included, as the early/late minima may not be found in the year of interest (the data range can be specified by the user). The same can be true for the peak. The user can select if a Northern hemisphere (peak assumed in the middle of the year), Southern hemisphere (peak is assumed around the turn of the year) or mixed phenology (both can happen, and FORCE TSA should attempt to decide for each pixel) should be assumed. In total, 26 LSP metrics will be derived for each year (see output subsection). All metrics will be output if LSP output is requested (this will give 26 physical files).

In a final step, time series analyses can be performed on the folded time series. The user can choose, which folded time series to use (daily, weekly, monthly or annual – folded with mean, min, max, or LSP). In case of annual LSP time series, the analysis is performed for each LSP metric (this will give you 26 physical files). Currently implemented analyses are linear trend analysis, and an extended CAT transform (with additional change and trend parameters for the three parts of the time series; see output subsection).

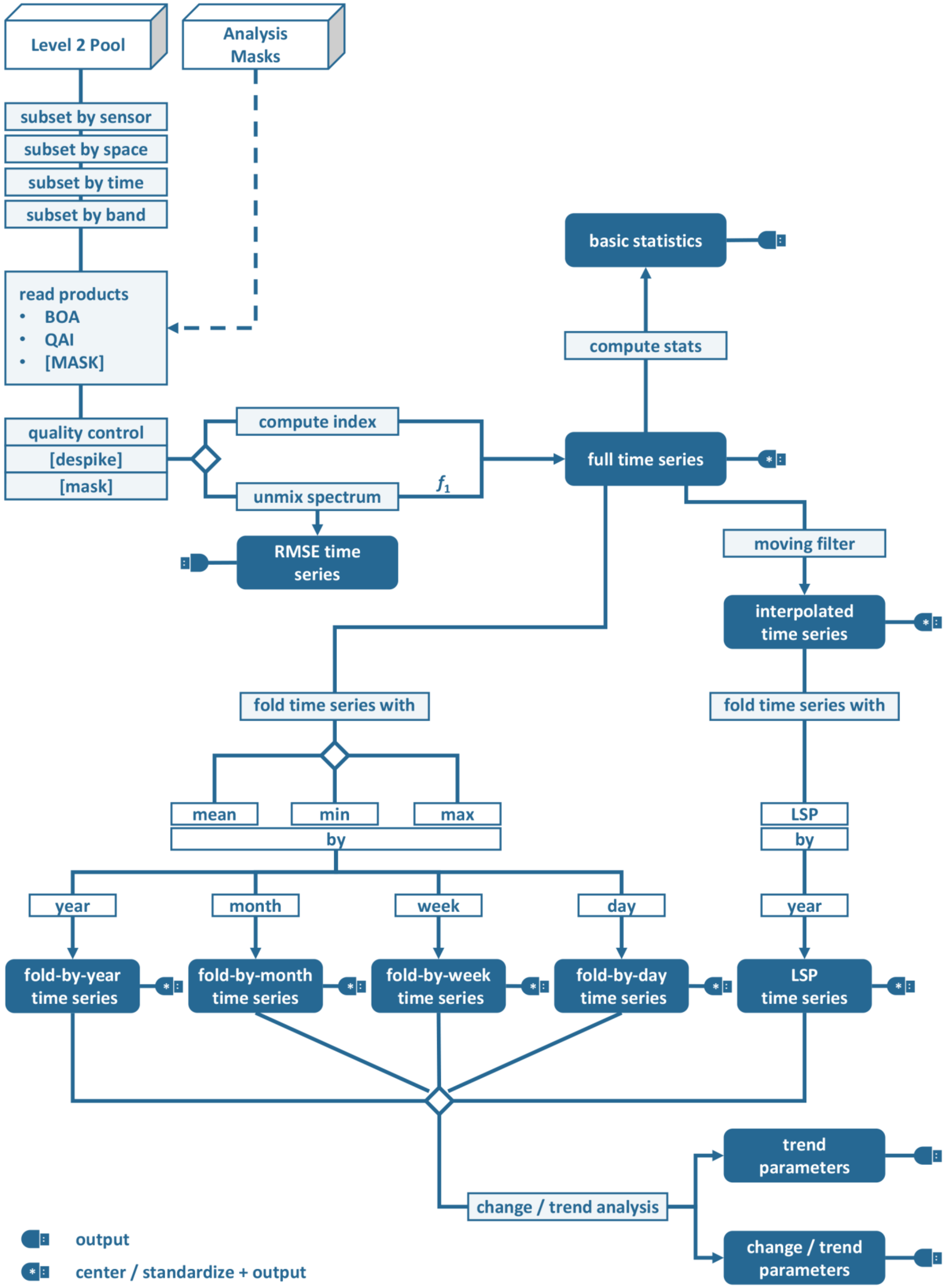

Fig. 5. Workflow of FORCE TSA.
The processing is done for each requested tile.

4) Analysis masks

A mask dataset may be input to constrain the analysis to a certain area or land cover. This is optional. If no masks are given, the full image extent is used. The mask dataset needs to be prepared in the same grid as the Level 2 data (i.e. in a mirrored data structure). The filenames must contain the keyword 'MASK'. The data are expected to be in ENVI standard or GeoTiff format with file extension *.dat/tif, *.hdr. Each file is a binary image with 1 for the 'on' pixels, 0 for the 'off' pixels.

5) Output format

(a) Data organization

The output data are organized in the gridded data structure used for Level 2 processing. The tiles manifest as directories in the file system, and the images are stored within. For mosaicking, use *force-mosaic* (will likely give an error for the TSS and RMS products because there are different numbers of observations in different tiles).

(b) Naming convention

Following 65-digit naming convention is applied to all output files:

```
1984-2017_182-274_LEVEL4_TSA_MULTI_TCG_C0_S0_FAVG_TY_C95T_TRD.tif
1984-2017_182-274_LEVEL4_TSA_MULTI_TCG_C0_S0_FAVG_TY_C95T_TRD.hdr
```

Digits 1–9 Temporal range for the years as YYYY–YYYY

Digits 11–17 Temporal range for the DOY as DDD–DDD

Digits 19–24 Product Level

LEVEL4	

Digits 26–28 Processing Type

TSA	

Digits 30–34 Set of spectral bands

LNDLG	Landsat legacy bands
SEN2L	Sentinel-2 land surface bands
SEN2H	Sentinel-2 high-res bands
R-G-B	Visual bands

Digits 36–38 3-digit index compact name

BLU	Blue band
GRN	Green band
RED	Red band
NIR	Near Infrared band
SW1	Shortwave Infrared band 1
SW2	Shortwave Infrared band 2
RE1	Red Edge band 1
RE2	Red Edge band 2
RE3	Red Edge band 3
BNR	Broad Near Infrared band

	NDV	Normalized Difference Vegetation Index
	EVI	Enhanced Vegetation Index
	NBR	Normalized Burn Ratio
	ARV	Atmospherically Resistant Vegetation Index
	SAV	Soil Adjusted Vegetation Index
	SRV	Soil and Atmospherically Resistant Vegetation Index
	TCB	Tasseled Cap Brightness
	TCG	Tasseled Cap Greenness
	TCW	Tasseled Cap Wetness
	TCD	Tasseled Cap-based Disturbance Index
	SMA	Spectral Mixture Analysis abundance
Digits 40–41	Center flag	
	C0	no
	C1	yes
Digits 43–44	Standardize flag	
	S0	no
	S1	yes
Digits 46–49	Folding method	
	FAVG	fold with average
	FMIN	fold with minimum
	FMAX	fold with maximum
	FLSP	fold with Land Surface Phenology
	F***	fold with Land Surface Phenology, in combination with TRD/CAT product type. Refer to next subsection.
Digits 51–52	Trend on …	
	TY	folded years
	TM	folded months
	TW	folded weeks
	TD	folded days
Digits 54–57	Significance parameters	
	C95T	95% (or other) confidence level, two-tailed t-test
	C95L	95% (or other) confidence level, left-tailed t-test
	C95R	95% (or other) confidence level, right-tailed t-test
Digits 59–61	Product Type	
	TSS	Time Series Stack
	RMS	RMSE Time Series of SMA
	STA	Basic Statistics
	TSI	Time Series Interpolation

TRD	Trend Analysis
CAT	Extended CAT Analysis
FBY	Fold-by-Year Stack
FBM	Fold-by-Month Stack
FBW	Fold-by-Week Stack
FBD	Fold-by-Day Stack
***	26 Land Surface Phenology metrics (refer to next subsection)

Digits 63–65 File extension

tif	image data in compressed GeoTiff format
dat	image data in flat binary ENVI format
hdr	metadata

(c) Special naming convention for Land Surface Phenology

Due to the one-to-26 nature of the LSP metrics, the naming convention is a bit more complex. First there are 26 LSP metrics, which are defined with following LSP name tags.

The outputted LSP metrics will have fold-by-LSP tag (digits 46–49: FLSP), and product type (digits 59–61) according to LSP name tag.

If Trend Analysis (TRD) or Extended CAT Analysis (CAT) product type based on LSP metrics is requested, the change / trend will be computed on each LSP metric. Therefore, naming convention needs to be modified. The outputted products will have folding method set to the LSP name tag (digits 46–49: e.g. FDEM, FVEM, …), and TRD / CAT product type (digits 59–61).

LSP name tags:

DEM	Date of Early Minimum
DSS	Date of Start of Season
DRI	Date of Rising Inflection
DPS	Date of Peak of Season
DFI	Date of Falling Inflection
DES	Date of End of Season
DLM	Date of Late Minimum
LTS	Length of Total Season
LGS	Length of Green Season
VEM	Value of Early Minimum
VSS	Value of Start of Season
VRI	Value of Rising Inflection
VPS	Value of Peak of Season
VFI	Value of Falling Inflection
VES	Value of End of Season
VLM	Value of Late Minimum
VBL	Value of Base Level
VSA	Value of Seasonal Amplitude

IST	Integral of Total Season
IBL	Integral of Base Level
IBT	Integral of Base+Total
IGS	Integral of Green Season
RAR	Rate of Average Rising
RAF	Rate of Average Falling
RMR	Rate of Maximum Rising
RMF	Rate of Maximum Falling

(d) Product type

⇒ *Time Series*

Time Series products have as many bands as there are available or requested time steps. If no temporal subset was specified:

the TSS product contains one band per available acquisition (this may vary between the tiles),

the RMS product contains one band per available acquisition (this may vary between the tiles),

the TSI product contains one band per interpolation step,

the FBY product contains one band per year (do not overdo YEAR_MIN/MAX, this will give many useless bands),

the FBM product contains one band per month (up to 12, depends on MONTH_MIN/MAX and DOY_MIN/MAX),

the FBW contains one band per week (up to 52, depends on MONTH_MIN/MAX and DOY_MIN/MAX),

the FBD product contains one band per DOY (up to 365, depends on MONTH_MIN/MAX and DOY_MIN/MAX),

the 26 LSP products contain one band per year (do not overdo YEAR_MIN/MAX, this will give many useless bands).

⇒ *Basic Statistics*

The Basic Statistics (STA) product provides a summary of all observations (or the requested subset). It is a multi-layer image with following bands:

1	μ	Average of index values
2	σ	Standard deviation of index values
3	min	Minimum index value
4	max	Maximum index value
5	# of obs.	Number of good quality observations

⇒ *Trend Analysis*

The Trend Analysis (TRD) product contains trend parameters. It is a multi-layer image with following bands:

1	μ	Average
2	a	Intercept
3	b	Trend
4	R^2	R squared
5	sig.	Significance (-1, 0, 1)
6	RMSE	Root Mean Squared Error
7	MAE	Mean Absolute Error

8	max |e|	Maximum Absolute Residual
9	# of obs.	Number of good quality observations

⇒ *Change, Aftereffect, Trend*

The Change, Aftereffect, Trend (CAT) product (following Hird et al. 2016, DOI: 10.1109/jstars.2015.2419594) contains extended change and trend parameters. It detects one change per time series, splits the time series into three parts, and derives trend parameters: (1) complete time series (this is the same as the TRD product), (2) time series before change, and (3) time series after change. It is a multi-layer image with following bands:

1	Change	Magnitude of change
2	Time of change	Timestamp of the change (depends on the input time series, i.e. year/month/week/day)
3–11	Trend parameters for complete time series (see TRD product)	
12–20	Trend parameters for time series before change (see TRD product)	
21–29	Trend parameters for time series after change (see TRD product)	

(e) File format

The data are provided in (i) ENVI Standard format (flat binary images), or (ii) as GeoTiff (LZW compression with horizontal differencing). Each dataset consists of an image dataset (.dat/,tif) and additional metadata (.hdr). The image data have signed 16bit datatype and band sequential (BSQ) interleaving. Scaling factor is 10000 for most products.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info, band names etc.

E. FORCE CSO – Clear Sky Observations

FORCE Clear Sky Observations (FORCE CSO) is intended for data availability mining. For given time steps (e.g. 3 months), per-pixel statistics about data availability are calculated, i.e. number of CSOs, and average (standard deviation, min, max, etc.) days between consecutive CSOs. A per-tile summary is written to the image header and a per-project summary is printed to screen. This section summarizes the usage of FORCE CSO, its helper programs and the output format.

FORCE CSO can only be used with Level 2 ARD.

A glimpse of what you get:

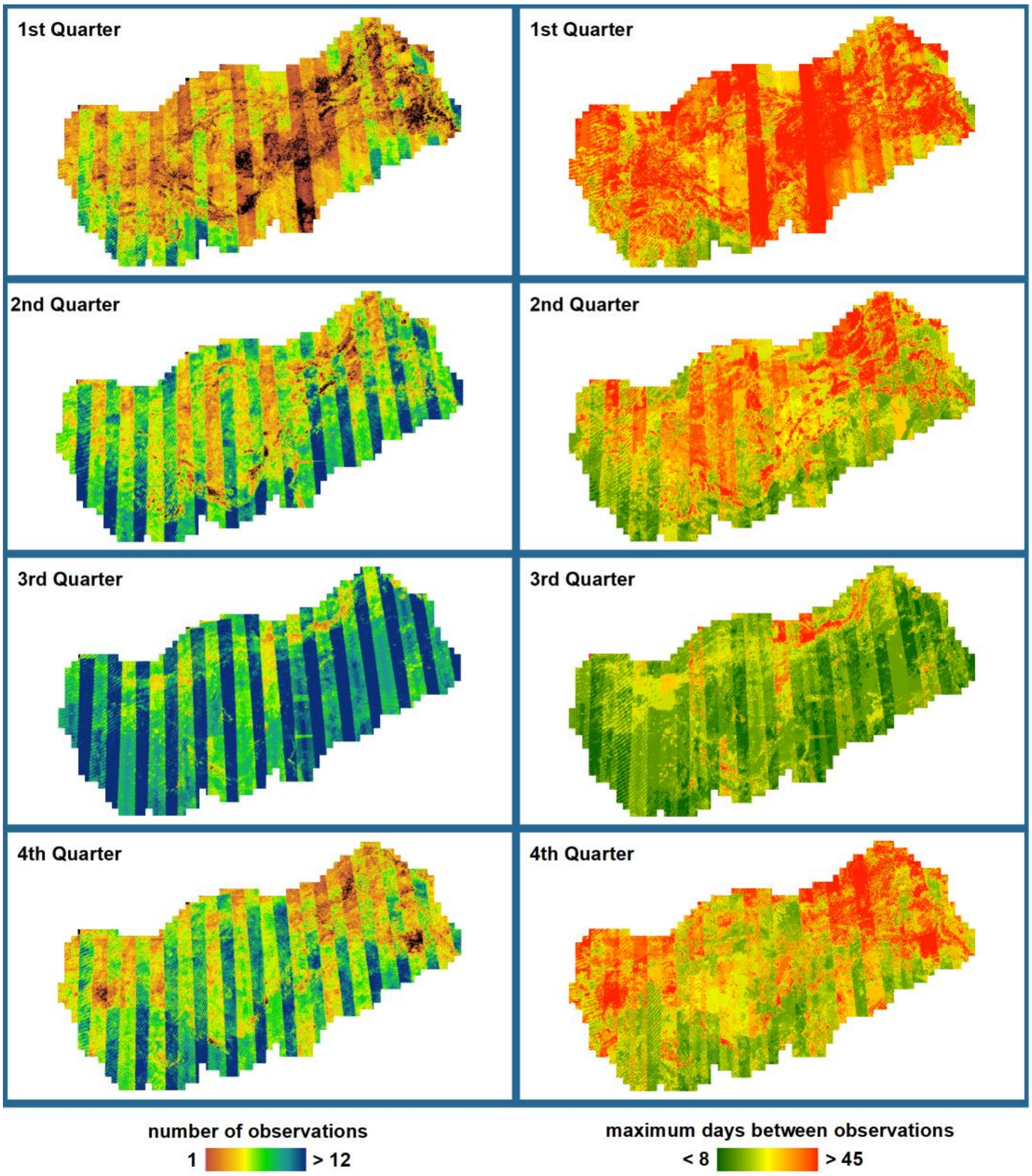

Fig. 6. Quarterly Clear Sky Observation statistics.
The CSO statistics were computed for the 2015 Landsat 7–8 acquisitions over Turkey.
[CSO statistics were generated with force-cso, then mosaicked with force-mosaic]

1) Parameterization

A parameter file is mandatory for FORCE CSO; *force-parameter-cso* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (1st argument). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.D summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-cso
Usage	|	force-parameter-cso dir

2) Processing

The core module of FORCE CSO is *force-cso*, which generates per-pixel statistics about data availability. The processed images are written to the output directory given in the CSO parameter file (see VII.D for all available options).

Module	|	force-cso
Usage	|	force-cso par-file

The parameter file needs to be given as 1st argument. All options need to be specified in the parameter file (see VII.D for all available options).

3) Quick guide

FORCE CSO is not covered by a scientific publication. Therefore, this section briefly summarizes the key functionality; see also section VII.G for all available options and possible keywords.

The analysis can be constrained to a certain area. The general extent in tile coordinates should be specified, and a tile white-list may be used.

The analysis can be performed with observations from all available sensors, or a subset of these. The spatial resolution on which the analysis is performed, needs to be given too. Image decimation/replication is taken care of (using nearest neighbor resampling).

Quality control is in full user control. All provided quality flags (see VI.B.5) can be used individually. Use this option rigorously!

The most important setting are the temporal properties. A temporal range needs to be specified in terms of years, e.g. 2017–2018. Clear Sky Observation statistics are generated for each time step as defined by the MONTH_STEP parameter (e.g. 6 months). Within each interval, the number of Clear Sky Observations are counted. In addition, several statistics based on the temporal difference between the observations are calculated, e.g. the maximum time between two observations in each interval. Note that the beginning and end of the intervals act as boundaries for this assessment and are also considered, for an example, see Fig. 7.

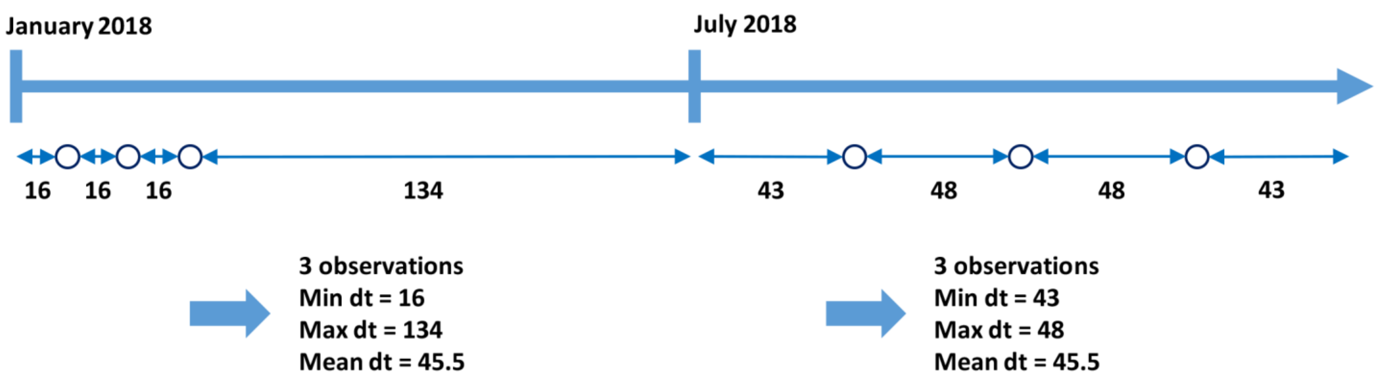

Fig. 7. Processing scheme of FORCE CSO.

This processing scheme reflects the fact, that a single measure of data availability might not yield representative results – depending on the application. A combined look at different statistics, or at a more uncommon metric, may provide more insight into the applicability of a specific method – or might explain uncertainties associated with a specific method. As an example, the data availability for the first and second half of 2018 as depicted in Fig. 7 is equal in terms of the number of observations, and the average time between observations. However, there are large differences in the maximum time between observations as the data are clumped in the first half. This may have important implications, e.g. for the detectability of harvesting events.

Currently available statistics are number of Clear Sky Observations, as well as the average, standard deviation, minimum, maximum, range, skewness, kurtosis, median, 25/75% quantiles, and interquartile range of the difference between CSOs. Happy data mining.

4) Output format

(a) Data organization

The output data are organized in the gridded data structure used for Level 2 processing. The tiles manifest as directories in the file system, and the images are stored within. For mosaicking, use *force-mosaic*.

(b) Naming convention

Following 37-digit naming convention is applied to all output files:

```
2000-2010_03M_CSO-STATS_LNDLG_NUM.tif
2000-2010_03M_CSO-STATS_LNDLG_NUM.hdr
```

Digits 1–9 Temporal range for the years as YYYY–YYYY

Digits 11–13 Temporal binning in months

Digits 15–23 Processing Type

CSO-STATS

Digits 25–29 Set of spectral bands

LNDLG	Landsat legacy bands
SEN2L	Sentinel-2 land surface bands
SEN2H	Sentinel-2 high-res bands
R-G-B	Visual bands

Digits 31–33 Product Type

NUM	Number of Observations (optional output, scale 1, nodata: 0)
AVG	Average of dt (optional output, scale: 1, nodata: 0)
STD	Standard Deviation of dt (optional output, scale: 1, nodata: 0)
MIN	Minimum of dt (optional output, scale: 1, nodata: 0)
MAX	Maximum of dt (optional output, scale: 1, nodata: 0)
RNG	Range of dt (optional output, scale: 1, nodata: 0)
SKW	Skewness of dt (optional output, scale: 10000, nodata: 0)
KRT	Kurtosis of dt (optional output, scale: 1000, nodata: 0)
Q25	0.25 Quantile of dt (optional output, scale: 1, nodata: 0)
Q50	0.50 Quantile of dt (optional output, scale: 1, nodata: 0)

	Q75	0.75 Quantile of dt (optional output, scale: 1, nodata: 0)
	IQR	Interquartile Range of dt (optional output, scale: 1, nodata: 0)
Digits 35–37	File extension	
	tif	image data in compressed GeoTiff format
	dat	image data in flat binary ENVI format
	hdr	metadata

(c) Product type

There are several product types available, all are optional. The Number of Observations (NUM) product contains the number of clear sky observations for each pixel and bin. The other products are statistical summaries of the temporal difference between consecutive observations within these bins.

(d) File format

The data are provided in compressed GeoTiff or flat binary ENVI Standard format. Each dataset consists of an image dataset (.tif/.dat) and metadata (.hdr). The image data have signed 16bit datatype and band sequential (BSQ) interleave format. The products have as many bands as there are intervals within the defined year range.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info, band names etc.

F. FORCE ImproPhe – Improving the Spatial Resolution of Land Surface Phenology

FORCE ImproPhe is intended to increase the spatial resolution of coarse continuous fields. It was originally developed to refine Land Surface Phenology metrics derived from MODIS, using sparse Landsat data as spectral and multi-temporal targets for data fusion. Regarding phenology, it can be used to obtain a Landsat-like phenology even in areas / during times when Landsat data alone is insufficient (in terms of temporal density). FORCE permits the use of Landsat and/or Sentinel-2 data as target datasets for the improPhement. ImproPhe can also be applied to other coarse resolution data (for best results, some link to spectral-temporal land surface processes should exist – e.g. increasing the spatial resolution of rainfall data won't work).

FORCE also features a tool to increase the spatial resolution of lower resolution ARD using higher resolution ARD using the ImproPhe algorithm (spectral + multi-temporal parameterization, see VI.G). The ImproPhe code is also implemented as resolution merge option in FORCE L2PS (VI.B) to increase the spatial resolution of Sentinel-2's 20m bands to 10m (spectral + mono-temporal implementation).

This section summarizes the usage of FORCE ImproPhe, its helper programs and the output format.

FORCE ImproPhe can only be used with Level 2 ARD.

A glimpse of what you get:

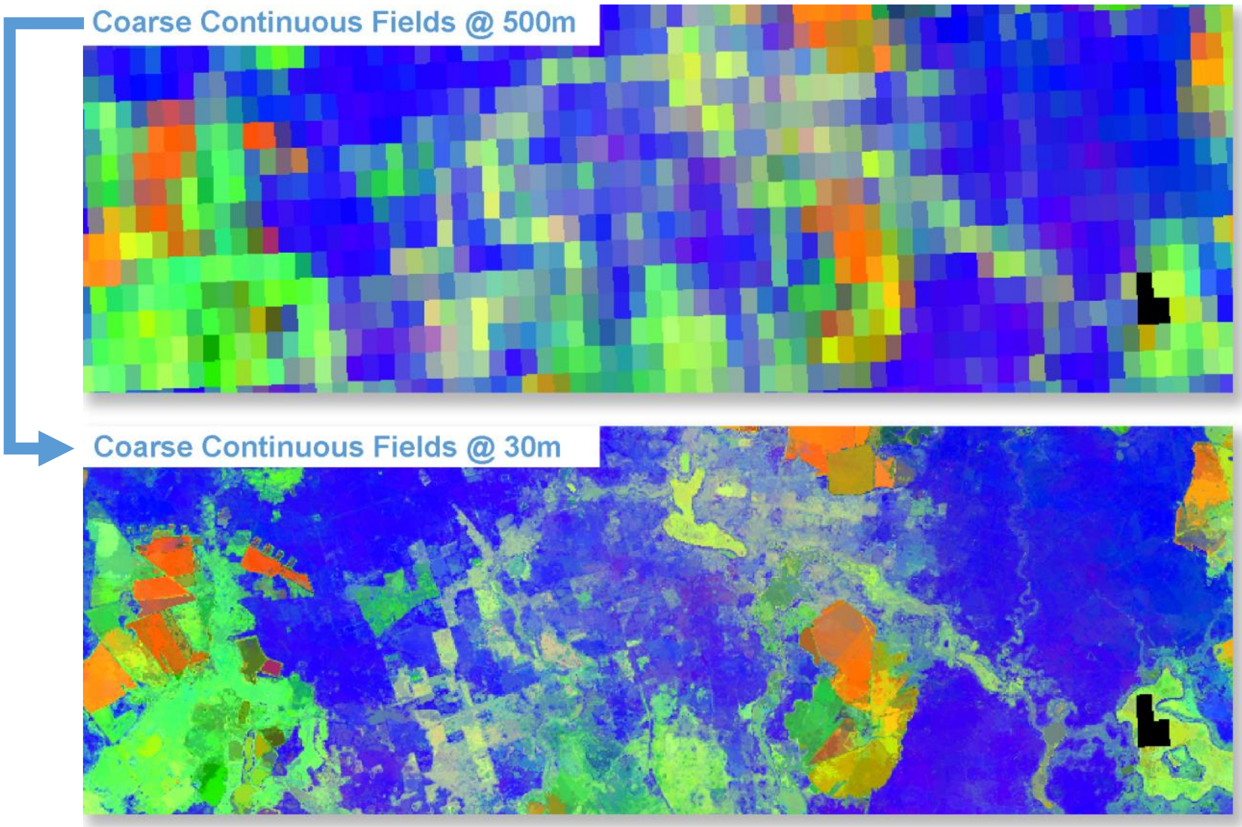

Fig. 8. Coarse resolution (500m) and ImproPhed (30m) LSP metrics.
Rate of Maximum Rise (R), Integral of Green Season (G), and Value of Early Minimum (B) phenometrics for a study site in Brandenburg, Germany. Using the ImproPhe algorithm, the LSP metrics were improved to 30m spatial resolution using Landsat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

1) Parameterization

A parameter file is mandatory for FORCE ImproPhe; *force-parameter-improphe* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (1st argument). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.H summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-improphe
Usage	|	force-parameter-improphe dir

2) Processing

The core module of FORCE ImproPhe is *force-improphe*, which improves the spatial resolution of coarse continuous fields. The processed images are written to the output directory given in the ImproPhe parameter file (see VII.H for all available options).

Module	|	force-improphe
Usage	|	force-improphe par-file

The parameter file needs to be given as 1st argument. All options need to be specified in the parameter file (see VII.H for all available options).

3) Output format

(a) Data organization

The output data are organized in the gridded data structure used for Level 2 processing. The tiles manifest as directories in the file system, and the images are stored within. For mosaicking, use *force-mosaic*.

(b) Naming convention

Following 21-digit naming convention is applied to all output files:

```
2017_IMPROPHE_IGS.tif
2017_IMPROPHE_IGS.tif
```

Digits 1–4	Year	
Digits 6–13	Processing Type	
	IMPROPHE	
Digits 15–17	Product Tag	
	XXX	These custom3-digit tags are specified in the parameter file
Digits 19–21	File extension	
	tif	image data in compressed GeoTiff format
	dat	image data in flat binary ENVI format
	hdr	metadata

(c) File format

The data are provided in compressed GeoTiff or flat binary ENVI Standard format. Each dataset consists of an image dataset (.tif/.dat) and metadata (.hdr). The image data have signed 16bit datatype. Each predicted image is stored as separate file.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info, band names etc.

G. FORCE L2IMP – Level 2 ImproPhe

FORCE Level 2 ImproPhe (L2IMP) is intended to increase the spatial resolution of lower resolution Level 2 ARD using higher resolution Level 2 ARD, e.g. to improve the spatial resolution of 30m Landsat imagery to 10m using Sentinel-2 data as targets. This only works for years where both data sources exist. The data fusion is performed with the ImproPhe algorithm. Note that this module is heavy on processing time.

This section summarizes the usage of FORCE L2IMP, its helper programs and the output format.

FORCE L2IMP can only be used with Level 2 ARD.

A glimpse of what you get:

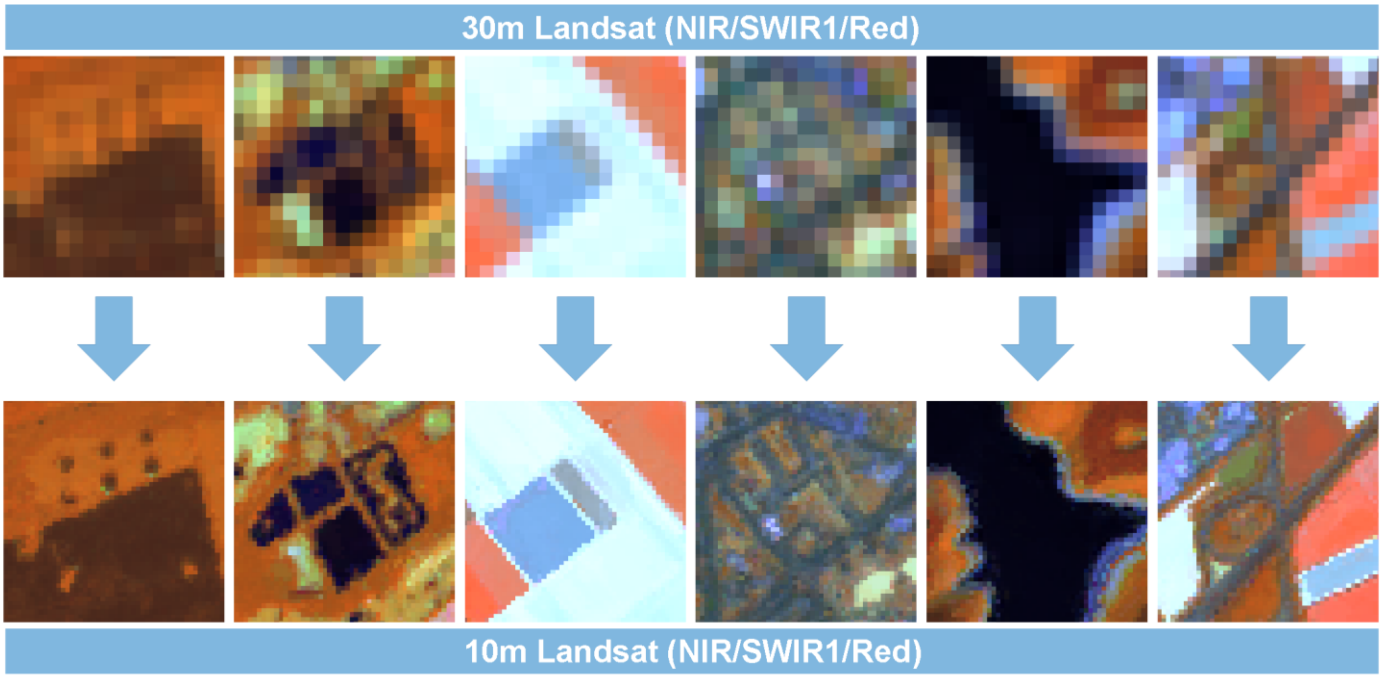

Fig. 9. 30m Landsat ARD, and ImproPhed 10m Landsat ARD.
The figure shows image subsets from North Rhine-Westphalia, Germany. Using the ImproPhe algorithm, the spatial resolution was improved to 10m using multi-temporal Sentinel-2 A/B high-res bands as prediction targets.
[Data were fused using force-l2imp]

1) Parameterization

A parameter file is mandatory for FORCE L2IMP; *force-parameter-l2imp* is a small helper that generates a parameter file skeleton, i.e. an empty parameter file. The skeleton is written to the specified directory (1st argument). The parameter file can (and should) be renamed and needs to be modified (with any text editor). Section VII.I summarizes the file format and a brief description of the parameters.

Module	|	force-parameter-l2imp
Usage	|	force-parameter- l2imp dir

2) Processing

The core module of FORCE L2IMP is *force- l2imp*, which improves the spatial resolution of coarser resolution ARD. The processed images are ingested into the Level 2 directory (see VII.I for all available options).

Module	|	force- l2imp
Usage	|	force- l2imp par-file

The parameter file needs to be given as 1st argument. All options need to be specified in the parameter file (see VII.I for all available options).

3) Output format

(a) Data organization

The output data are organized as the Level 2 data. The output data are appended to the input Level 2 data as new product, i.e. two additional files (image + metadata) appear next to the existing data. Note that the new product will have higher spatial resolution (more pixels) than the other products (e.g. QAI). Higher-level FORCE routines can handle this. For any higher-level FORCE module, you can choose to use the improved product or the original one.

(b) Naming convention

Following 29-digit naming convention is applied to all output files:

```
20180823_LEVEL2_LND08_IMP.tif
20180823_LEVEL2_LND08_IMP.hdr
```

The naming convention is the same as for the Level 2 data (see VI.B.5). The only difference is the Product Type, which is set to IMP.

Digits 23–25 Product Type

IMP ImproPhed Bottom-of-Atmosphere Reflectance (standard output, scale: 10000, nodata: -9999)

(c) Product type

There is only one product type, i.e. the ImproPhed Bottom-of-Atmosphere Reflectance (IMP). The IMP product has the same specification as the BOA product, but spatial resolution was enhanced.

(d) File format

The data are provided in compressed GeoTiff or flat binary ENVI Standard format. Each dataset consists of an image dataset (.tif/.dat) and metadata (.hdr). The image data have signed 16bit datatype. Each predicted image is stored as separate file.

The metadata (.hdr) are provided in ENVI Standard format as human-readable text using tag and value notation. Metadata include image characteristics like dimensions, data type, band interleave, coordinate reference system, map info, band names etc.

H. FORCE WVDB - Water Vapor Database

The FORCE Water Vapor Database (FORCE WVDB) component can be used to generate and maintain a water vapor database used for atmospheric correction of Landsat data (within FORCE L2PS). Instead of running this component on your own, an application-ready global water vapor database (2000–July 2018) can be downloaded from https://doi.org/10.1594/PANGAEA.893109. The water vapor database is not used when processing Sentinel-2 data because water vapor can be estimated from the images themselves.

A glimpse of what you get:

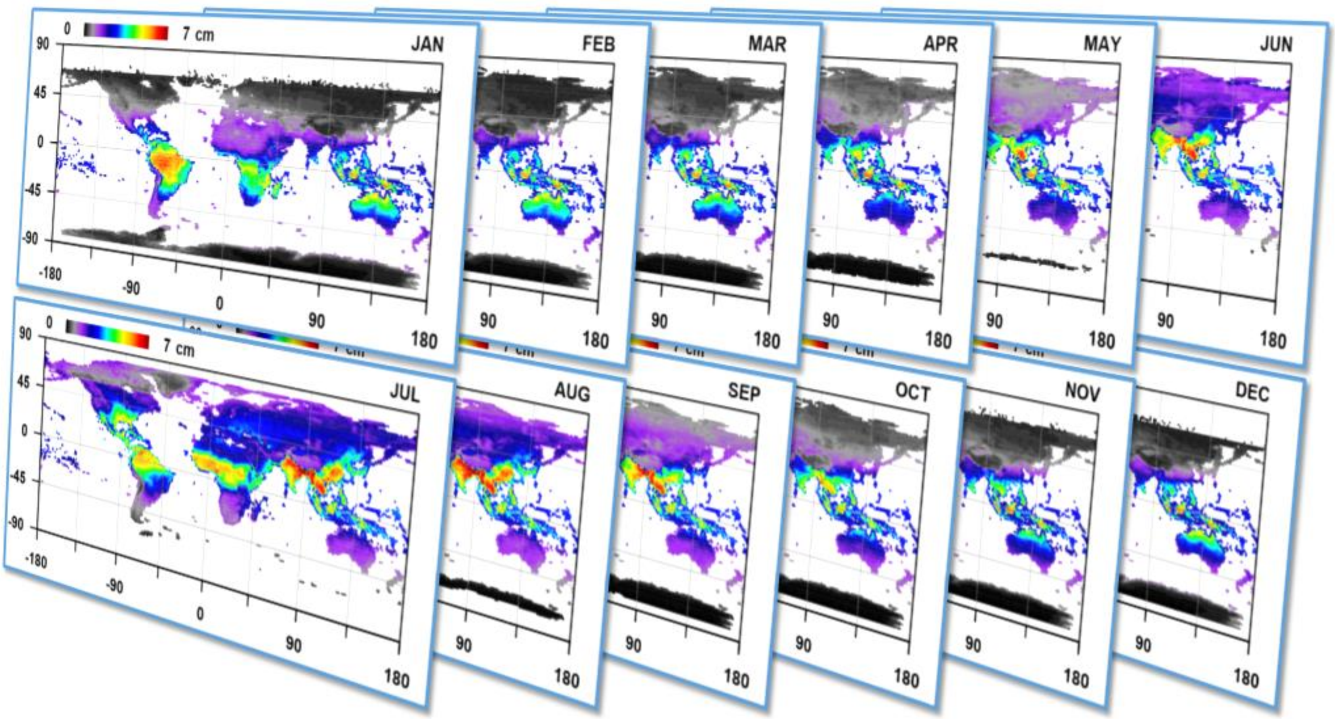

Fig. 10. Water vapor climatology.
Global, and monthly water vapor climatology for each land-intersecting Landsat WRS-2 scene.
[The daily water vapor database, and the climatology were generated with force-lut-modis.
This dataset is freely available here: https://doi.org/10.1594/PANGAEA.893109]

1) Generation

The water vapor database may be generated/updated with *force-lut-modis*, which automatically downloads MODIS water vapor products (MOD05 and MYD05, collection 6.1) from the Level1 and Atmosphere Archive and Distribution System (LAADS) at NASA's Goddard Space Flight Center. Note that any permanent or temporary change/shutdown/decommissioning on LAADS' or MODIS' end may result in the nonfunctioning of *force-lut-modis* (as it happened between the release of FORCE v. 1.1 and v. 2.0). The data are processed to supply daily and climatological water vapor averages per given coordinate (± 0.5°).

The usage of this tool is not mandatory, but water vapor tables are needed for the radiometric processing of Landsat data (unless one single dummy value is specified in the Level 2 parameter file). However, the tables can also be generated with other data and/or software. If you want to create the tables on your own, refer to the file format description (see section VII.J).

An application-ready global water vapor database (daily values for 2000–July 2018 + monthly climatology) can be downloaded from https://doi.org/10.1594/PANGAEA.893109.

Module	|	force-lut-modis
Usage	|	force-lut-modis coords dir-wvp dir-geometa dir-eoshdf [start-year start-month start-day] [end-year end-month end-day]

The first four arguments are mandatory, the start and end arguments are optional and may be used for parallelization during the initial build.

A text file with coordinates needs to be given as 1st argument. Water vapor averages are estimated for each given coordinate. The file should contain two columns separated by white space, and no header. The first column should give the longitude, the second column the latitude with coordinates in decimal degree (negative values for West/South), see section VII.J.

The MODIS data are downloaded to dir-eoshdf (this directory must exist). MODIS data that are already in dir-eoshdf are not downloaded again. If the tool crashes because a dataset is corrupt, it is necessary to manually delete this file and run the tool again. Unfortunately, this happens from time to time due to incomplete downloads or if LAADS is unresponsive. The program attempts to re-download a corrupt file up to 10 times, but this error can occur nonetheless.

MOD05/MYD05 data are swath products, and MOD03/MYD03 geometa tables are necessary to relate coordinates to MODIS granules. The geometa tables are downloaded to dir-geometa (this directory must exist). Tables that are already in dir-geometa are not downloaded again. If the tool crashes because a table is invalid, it is necessary to manually delete this file and run the tool again. Unfortunately, this happens from time to time due to incomplete downloads or if LAADS is unresponsive. The program attempts to re-download a corrupt file up to 10 times, but this error can occur nonetheless. Note that the geometa tables are a global product, and as such, the same dir-geometa can be used for different projects (≙ study areas).

The final water vapor tables are saved in dir-wvp (this directory must exist). Tables that are already in dir-wvp are not processed again (i.e. no download of geometa tables and hdf files; the user may delete the MODIS *.hdf files after the water vapor tables are successfully generated.). This directory is the directory, to which DIR_WVPLUT in the Level 2 parameter file (see section VII.A) should refer.

MODIS data are not available for each day (pre-MODIS era, sensor outages, clouds…). Therefore, we use a climatology as fallback option. The climatology is the monthly long-time average of the daily values. The climatology is generated after processing the daily values. It is strongly recommended to re-run the tool – in sequential mode – after all the daily tables were successfully generated (to have a clean version of the climatology tables). This is especially mandatory if the tool was run in parallel during the initial build.

The initial build of the database might take some time – strongly depends on the size of the study area. This is mainly due to the download of MODIS data[1]. It is possible to speed up the work by parallel calls to *force-lut-modis*. However, note that LAADS restricts the number of running downloads and may permanently block your IP address (LAADS FAQ). The optional start and end dates may be used for parallelization, e.g. one processor per year.

Example for the initial build using *GNU parallel*:

```
seq -w 2000 2017 | parallel -j18 force-lut-modis coords dir-wvp dir-geometa dir-eoshdf {} 01 01 {} 12 31
```

[1] It is also possible to acquire all necessary files in any other way, and place them in dir-eoshdf. Existing files will not be downloaded.

I. FORCE AUX - Auxiliary Functionality

The FORCE Auxiliary Functionality (FORCE AUX) component is intended to provide small helper programs for specific purposes, e.g. the location of coordinates.

A glimpse of what you get:

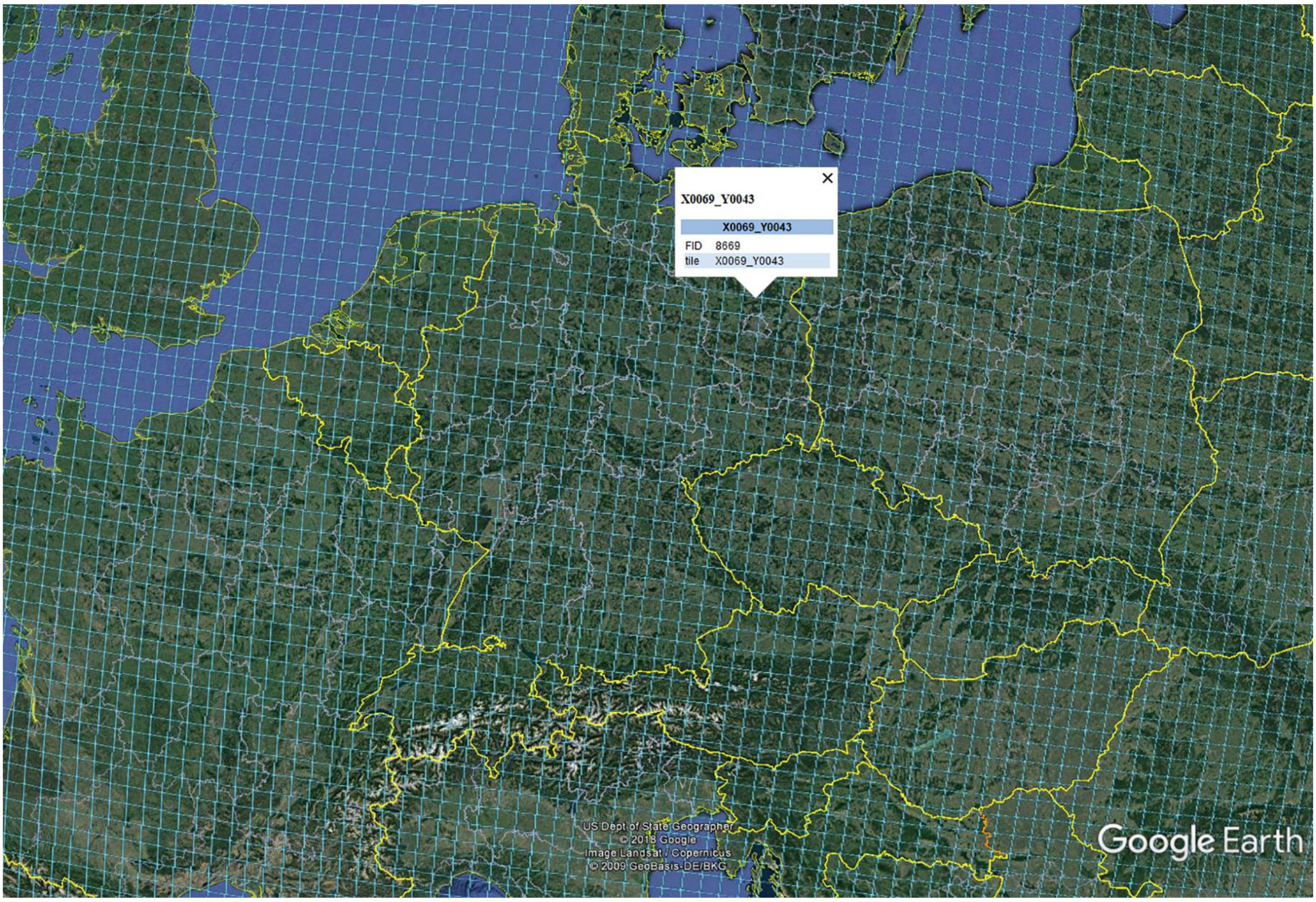

Fig. 11. ARD processing grid.
This figure shows a Google Earth screenshot with ARD grid overlay. The grid cells are 30 x 30km in EPSG:3035.
[The grid was generated using force-tabulate-grid]

1) Print version

The version of your FORCE installation, as well as a short disclaimer can be displayed using the utility program *force*.

Module	force
Usage	force

2) Inflate QAI bit layers

Quality Assurance Information (QAI) is generated for Level 2 data (QAI product), and is included as layer in the Level 3 compositing information product (INF). The QAI layers are stored with bit-encoding (see Table 6), which makes it very useful to store much information with fairly low data volume. However, the QAI need to be parsed to extract all the useful information. The program *force-qai-inflate* can be used to inflate the QAI to individual masks. The 1st argument specifies the QAI dataset that should be inflated (QAI or INF products for Level 2 and Level 3, respectively). The 2nd argument is the directory where the masks should be stored. The 3rd argument gives the output format (ENVI or GTiff). The output is a multilayer image (bands = flags in Table 6) with product type QIM (Quality Inflated Masks). It is advised to not store these images in the same directory as the QAI image. It is also advised to not use this tool operationally, adjust your programs to use the QAI layer directly (saves disk space and time).

Module	force-qai-inflate
Usage	force-qai-inflate QAI dir format

3) Coordinate locator

In case of the gridded data structure, a geographic coordinate can be converted to tile and pixel coordinates using *force-tile-finder*. This functionality is intended to spatially locate data. Any gridded data cube (containing a data cube definition file, see VII.M) can be given as 1st argument. Longitude and latitude must be given as 2nd and 3rd arguments with coordinates in decimal degree (negative values for West/South). The resolution needs to be given (4th argument) in order to relate coordinates to pixel positions.

Module	force-tile-finder
Usage	force-tile-finder datacube lon lat res

4) Export grid

In case of the gridded data structure, *force-tabulate-grid* can be used to extract the processing grid as ESRI shapefile, e.g. for visualization purposes or to generate a tile white-list. Any gridded data cube (containing a data cube definition file, see VII.M) can be given as 1st argument. The approximate bounding box of your study area needs to be given with coordinates in decimal degree (negative values for West/South). The shapefile 'datacube-grid.shp' is stored in the same directory as the data cube.

Module	force-tabulate-grid
Usage	force-tabulate-grid datacube bottom top left right

VII. AUXILIARY FILES

A number of auxiliary files are needed (or are optional) and must be prepared according to the specification detailed in the corresponding subsections. Table 7 summarizes all files with a short description of their purpose. All files are plain text files and may be edited and generated with any text editor. **The End-of-Line character must be in UNIX format**, i.e. line feed (\n); if there is a reading error or any arbitrary error, check EOL.

Table 7. FORCE auxiliary files.

AUX file	Short description	used by	mandatory	Page
Level 2 parameter file	Parameterization of Level 2 processing	*force-level2* *force-l2ps* *force-tile-finder* *force-tabulate-grid* ► see *force-parameter-level2*	yes yes yes yes	59
Level 2 file queue	File queues for Level 2 batch-processing	*force-level2* ► see *force-level1-landsat* ► see *force-level1-sentinel2*	yes	63
Tile white-list	White-listing of specific tiles that should be processed / output	*force-level2* *force-l2ps* *force-level3* *force-tsa* *force-cso* *force-improphe* *force-l2imp*	no no no no no no no	64
Level 3 parameter file	Parameterization of Level 3 processing	*force-level3* ► see *force-parameter-level3*	yes	65
TSA parameter file	Parameterization of Time Series Generation / Analysis	*force-tsa* ► see *force-parameter-tsa*	yes	70
Endmember file	Endmember file for spectral unmixing	*force-tsa*	no	75
CSO parameter file	Parameterization of Clear Sky Observation data mining	*force-cso* ► see *force-parameter-cso*	yes	76
ImproPhe parameter file	Parameterization of ImproPhe processing	*force-improphe* ► see *force-parameter-improphe*	yes	78
L2IMP parameter file	Parameterization of Level 2 ImproPhe processing	*force-l2imp* ► see *force-parameter-l2imp*	yes	81
Water vapor LUTs	Look-up-Tables for using external water vapor values in Level 2 Landsat processing	*force-level2* *force-l2ps* ► see *force-lut-modis*	no no	84
AOD LUTs	Look-up-Tables for using external / fallback AOD values in Level 2 processing	*force-level2* *force-l2ps*	no no	85
Coordinate file	Definition of coordinates that should be processed with some functionality	*force-lut-modis*	yes	85
Data Cube defintion	The spatial definition of a data cube.	*all* ► *force-l2ps + following modules*	yes	85

► files can be generated using the indicated programs

A. Level 2 parameter file

The Level 2 parameter file is mandatory for FORCE L2PS. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_LEVEL2_START++ and ++PARAM_LEVEL2_END++ keywords enclose the parameter file. Most parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The PROJECTION parameter is an exception and needs to be specified in two lines (because it contains '=' and spaces). The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-level2*.

1) Files and directories

All files and directories must exist and must be readable. DIR_LEVEL2, DIR_TEMP and DIR_DARKOB must also be writeable. Avoid relative filenames. All values in this section are interpreted as characters. Rudimentary checks are performed by the software and you will be notified if parameters are missing, invalid or if a specific combination does not work. Note that these checks might not be exhaustive.

`FILE_QUEUE`

The file queue specifies which images are to be processed. The full path to the file needs to be given (do not paste the content of the file queue into the parameter file). It is mandatory for *force-level2* but may be NULL for *force-l2ps*. See section VII.B for the file format.

`DIR_LEVEL2`

This is the output directory where the Level 2 data will be stored. Note that data will be overwritten/mosaicked if you reprocess images. The logfile generated by force-level2 will also appear in this directory. It is safe and recommended to use a single Level 2 data pool for different sensors (provided the same grid and projection is used). The higher-level programs of FORCE can handle different spatial resolutions (e.g. 30 m Landsat and 10 m Sentinel-2).

`DIR_TEMP`

This is a temporary directory that is used to extract the Landsat tarballs in *force-level2*. The extracted data will be deleted once they were processed. If you cancel processing, you may want to delete any left-overs in this directory.

A file 'cpu-$TIME' is temporarily created in DIR_TEMP. This file can be modified to re-adjust the number of CPUs while(!) *force-level2* is running. Note that the effect is not immediate, as the load is only adjusted after one of the running jobs (images) is finished.

`DIR_WVPLUT`

This is the directory where the water vapor tables are located (see section VII.J). Water vapor tables may be generated using FORCE WVDB (VI.H) or can be downloaded from https://doi.org/10.1594/PANGAEA.893109.

Water vapor tables are not required for Sentinel-2, in this case DIR_WVPLUT may be NULL. For Landsat, it is recommended to use this functionality.

As a minimum requirement, DIR_WVPLUT may be NULL and a global value for WATER_VAPOR (see below) needs to be specified.

If a directory is given, WATER_VAPOR is ignored. DIR_WVPLUT must contain water vapor tables. The 12 climatology tables must exist at least. They are used if the daily tables do not exist or if there is no valid daily value. FORCE WVDB can be used to build theses tables from MODIS data. The tables can also be generated using any other software/data; in this case refer to the table structure (see section VII.J).

If DO_ATMO = FALSE, DIR_WVPLUT and WATER_VAPOR are ignored.

`DIR_AOD`

This is the directory where the aerosol optical depth look-up-tables are located (see section VII.K). They can be used to input external AOD values. We commonly do not use this option. The usage of these values depends on DO_ATMO and DO_AOD (i.e. estimate AOD from the image, see below).

If DO_ATMO = FALSE, all AOD-related parameters are ignored.

If DO_ATMO = TRUE and DO_AOD = FALSE, DIR_AOD must be given and the external values are used.

If DO_ATMO = TRUE and DO_AOD = TRUE and DIR_AOD = NULL, AOD is estimated from the image. If there are no valid dark targets, AOD is set to a global surrogate value. This is the recommended setting.

If DO_ATMO = TRUE and DO_AOD = TRUE and DIR_AOD is given, AOD is estimated from the image. If there are no valid dark targets, the external AOD values are used as fallback option.

`FILE_DEM`

This file specifies the DEM mosaic (see VI.B.2). It is highly recommended to use a DEM. It is used for cloud / cloud shadow detection, atmospheric correction and topographic correction. The user can choose to process without a DEM (FILE_DEM =

NULL); in this case the surface is assumed flat @ z = 0m. Topographic correction cannot be used without a DEM. The quality of atmospheric correction and cloud /cloud shadow detection will suffer without DEM.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be output. The usage of FILE_TILE depends on DO_TILE.
If DO_TILE = FALSE, the parameter is ignored and can be NULL.
If DO_TILE = TRUE, all tiles can be output (NULL) or output restrictions can be used (specify a tile white list).

2) Spatial properties

`DO_REPROJ`

This indicates whether the images should be reprojected to the target coordinate system or if they should stay in their original UTM projection. This parameter is interpreted as logical value. For ARD generation, you should reproject.

`DO_TILE`

This indicates whether the images should be gridded after processing. This parameter is interpreted as logical value.
Depending on DO_TILE, sub-directories for the tiles (TRUE) or the original spatial reference systems (FALSE) will be created. For ARD generation, you should tile the data.

`TILE_SIZE`

This is the tile size (in target units, commonly in meters) of the gridded output; tiles are square; not used if DO_TILE = FALSE. This parameter is interpreted as float value.
It is advised to use a reasonable compromise that enables spatial coverage, and small tile sizes that increase input speed for higher-level functionality. We currently recommend 30,000m.
Note that some FORCE modules (FORCE ImproPhe, L2IMP) use data from the neighboring 8 tiles. Do not use tile sizes smaller than the typical prediction radii (see VII.H, and VII.I).

`RESOLUTION`

This is the spatial resolution of the output; not used if DO_REPROJ = FALSE. This parameter is interpreted as float value. We recommend 30m for Landsat, and 10m for Sentinel-2.

`ORIGIN LAT / ORIGIN LON`

These are the origin coordinates of the grid system (see Fig. 12) in decimal degree (negative values for West/South). The upper left corner of tile X0000_Y0000 represents this point. It is a good choice to use a coordinate that is North-West of your study area – to avoid negative tile numbers. Not used if DO_TILE = FALSE. These parameters are interpreted as float values.
Known problem: if a UTM projection is chosen, the origin point should be located relatively close to the study area – the *GDAL* coordinate conversion routine will fail otherwise and the code will crash.

`PROJECTION`

This defines the target coordinate system. The coordinate system must be given as WKT string. Note that – due to technical reasons – the string must be placed in the line following the PROJECTION tag. If DO_REPROJ = FALSE, the projection string can be NULL. We recommend to use equal area projections.

`RESAMPLING`

This is the resampling option for the reprojection; you can choose between Nearest Neighbor (NN), Bilinear (BL) and Cubic Convolution (CC); not used if DO_REPROJ = FALSE. This parameter is interpreted as string.

`RES_MERGE`

This defines the method used for improving the spatial resolution of Sentinel-2's 20 m bands to 10 m. Pixels flagged as cloud or shadow will be skipped.
The default algorithm is IMPROPHE, which uses the ImproPhe code in a spectral-only setup. A multi-parameter regression prediction can be used (REGRESSION); results are expected to be best, but processing time is significant. Another option is to use a spectral-only setup of the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM), although prediction artifacts occur between land cover boundaries. NONE disables resolution merge; in this case, 20m bands are quadrupled.

`TIER`

This specifies the acceptable tier level. In the case of Landsat, TIER = 1 will only accept L1T images; TIER = 2 will also accept L1Gt and L1G images. This parameter is interpreted as logical value.

3) Radiometric correction options

`DO_TOPO`

This indicates whether topographic correction should be performed. This parameter is interpreted as logical value. DO_TOPO can only be used if DO_ATMO = TRUE – and if a DEM is given.

`DO_ATMO`

This indicates whether atmospheric correction should be performed (FALSE: TOA reflectance, TRUE: BOA reflectance). This parameter is interpreted as logical value.

`DO_BRDF`

This indicates whether a bulk correction for bi-directional effects should be used. If TRUE, the output is nadir BRDF adjusted reflectance instead of BOA reflectance (the output is named BOA nonetheless). This parameter is interpreted as logical value.

If DO_ATMO = FALSE, DO_BRDF is forced to be FALSE.

`DO_AOD`

This indicates whether the internal AOD estimation (TRUE) or externally generated AOD values should be used (see DIR_AOD). This can also be parameterized as fallback option. This parameter is interpreted as logical value.

4) Advanced Radiometric correction options

`MULTI_SCATTERING`

This indicates whether the multiple scattering (TRUE) or the single scattering approximation (FALSE) should be used. This parameter is interpreted as logical value.

`ADJACENCY EFFECT`

This indicates whether to correct for the adjacency effect. This parameter is interpreted as logical value.

`SUN_VIEW_GRID`

This is the grid size (in meters) for which the angular parameters are calculated, i.e. sun position and viewing geometry. Note that this parameter has a large impact on the processing time if a very small value is used. The maximum number of grid cells can be 65,534. We suggest to use 5 km, in consistency with the Sentinel-2 Level 1C data product. This parameter is interpreted as integer value.

`WATER_VAPOR`

This specifies a global value for atmospheric water vapor content in g cm^{-2}. This parameter is interpreted as float value. This parameter can be a dummy value to quickly process an image without needing to generate a water vapor database. Note that especially Landsat-8 is relatively insensitive to atmospheric water vapor, and external water vapor is not needed to process Sentinel-2. The error in using a dummy value is significant for the TM sensors.

This value is only used if DIR_WVPLUT = NULL and DO_ATMO = TRUE.

5) Cloud detection options

`MAX_CLOUD_COVER_FRAME`

This is a threshold which suppresses the output for images that exceed the given cloud coverage. Note that images with 100% cloud coverage won't be output due to algorithmic reasons. This parameter is interpreted as float value.

`MAX_CLOUD_COVER_TILE`

This works on a tile basis. If the cloud coverage within a chip exceeds this threshold, the output is suppressed. Note that images with 100% cloud coverage won't be output due to algorithmic reasons. This parameter is interpreted as float value.

`CLOUD_THRESHOLD and SHADOW_THRESHOLD`

These are the main thresholds of Fmask. Default values are 0.225 and 0.02, respectively. These parameters are interpreted as float values.

6) Output options

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

`OUTPUT_CLD`

This parameter controls whether the cloud and cloud distance product should be output. Valid values are TRUE / FALSE. Use TRUE, if you want to build composites. Note that this is not the actual cloud mask. The cloud distance gives the distance to the next opaque cloud, buffered cloud, cirrus cloud, cloud shadow or snow. For cloud masks and quality screening, rather use the QAI product.

`OUTPUT_AOD`

This parameter controls whether the aerosol optical depth (@ 550 nm) product should be output. Valid values are TRUE / FALSE.

`OUTPUT_WVP`

This parameter controls whether the water vapor estimate product should be output. Valid values are TRUE / FALSE.

`OUTPUT_VZN`

This parameter controls whether the view zenith angle product should be output. Valid values are TRUE / FALSE. Use TRUE, if you want to build composites.

`OUTPUT_HOT`

This parameter controls whether the Haze Optimized Transformation product should be output. Valid values are TRUE / FALSE. Use TRUE, if you want to build composites.

B. *Level 2 file queue*

This file queue is mandatory for *force-level2*. It specifies the input images that are to be processed. One image per line should be given. The full file paths must be given. No white spaces should be present in the file paths. The file is specified with FILE_QUEUE in the Level 2 parameter file (see section VII.A).

Each image is followed by 'QUEUED' or 'DONE', which indicates the queue status. Queued images will be processed, and the queue status will be changed to 'DONE' after Level 2 processing. If a reprocessing is required, the queue status needs to be changed to 'QUEUED', e.g. using

```
sed -i 's/DONE/QUEUED/' level1-landsat-germany.pool
```

File queues can be generated – and updated with new acquisitions – using the FORCE L1AS programs (see section VI.A).

Although not specifically required, we recommend to use a consistent and clean data pool that contains all (and nothing else) input images. Note that images downloaded from space agencies may be redundant, i.e. multiple instances of the same file with different processing IDs are possible (thus new data may be overwritten by outdated data in the Level 2 data pool). A clean data pool can be generated and maintained with FORCE L1AS.

Example for Landsat:

```
/data/africa/miombo/level1/landsat/177072/LT51770722008065JSA00.tar.gz QUEUED
/data/africa/miombo/level1/landsat/177072/LC81770722014129LGN00.tar.gz QUEUED
/data/africa/miombo/level1/landsat/173070/LE71730701999276SGS00.tar.gz QUEUED
/data/africa/miombo/level1/landsat/173070/LC81730702014213LGN00.tar.gz QUEUED
… [file truncated]
```

Example for Sentinel-2:

```
/data/africa/miombo/level1/sentinel/T33LYD/S2A_MSIL1C_20170706T083601_N0205_R064_
   T33LYD_20170706T090107.SAFE/GRANULE/L1C_T33LYD_A010643_20170706T090107 QUEUED
/data/africa/miombo/level1/sentinel/T33LYC/S2A_MSIL1C_20170706T083601_N0205_R064_
   T33LYC_20170706T090107.SAFE/GRANULE/L1C_T33LYC_A010643_20170706T090107 QUEUED
/data/africa/miombo/level1/sentinel/T33LZE/S2A_MSIL1C_20170706T083601_N0205_R064_
   T33LZE_20170706T090107.SAFE/GRANULE/L1C_T33LZE_A010643_20170706T090107 QUEUED
/data/africa/miombo/level1/sentinel/T33LZF/S2A_MSIL1C_20170706T083601_N0205_R064_
   T33LZF_20170706T090107.SAFE/GRANULE/L1C_T33LZF_A010643_20170706T090107 QUEUED
… [file truncated]
```

C. *Tile white-list*

This file is optional, and may be used to suppress the output of tiles that are not in your study area. This option was implemented to decrease the volume of the processed data. The file extension is '.til'. The file is specified with FILE_TILE in the various parameter files. The file must be prepared as follows: the 1st line must give the number of tiles for which output should be created. The corresponding tile IDs must be given in the following lines, one ID per line; end with an empty line. The sorting does not matter.

The tool *force-tabulate-grid* can be used to create a shapefile of the grid, which in turn can be used to determine the desired tiles (see section VI.I.4).

Example:

```
4524
X0044_Y0014
X0044_Y0015
X0045_Y0013
X0045_Y0014
X0045_Y0015
X0045_Y0016
X0045_Y0017
X0045_Y0018
… [file truncated]
```

D. *Level 3 parameter file*

The Level 3 parameter file is mandatory for FORCE L3PS. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_LEVEL3_START++ and ++PARAM_LEVEL3_END++ keywords enclose the parameter file. All parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-level3*.

1) Directories

`DIR_LEVEL2`

This is the input directory, where the Level 2 ARD are stored.

`DIR_LSP`

This is the directory, where the Land Surface Phenology dataset is stored. This parameter may be NULL, if the phenology-adaptive method is disabled, i.e. DO_PHEN = FALSE.

`DIR_LEVEL3`

This is the output directory, where the Level 3 data will be stored.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be processed/output. If FILE_TILE = NULL, all available tiles (within the spatial extent; see below) are output.

2) Sensor white list

`SENSORS`

This parameter expects a list of sensor IDs that should be used for compositing. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = LND07 LND08 SEN2A SEN2B
```

Single-sensor composites/statistics have the same bands as the corresponding Level 2 data. Multi-sensor composites only contain overlapping bands. Exclusive bands are discarded. Note that no spectral adjustment is made. Following sensors are available:

LND04 (6-band Landsat 4 TM), LND05 (6-band Landsat 5 TM), LND07 (6-band Landsat 7 ETM+), LND08 (6-band Landsat 8 OLI), SEN2A (10-band Sentinel-2A), SEN2B (10-band Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). Following matching table (Table 8) is applied to determine overlapping bands:

Table 8. Sensor matching table for higher-level FORCE processing.
The numbers are the band numbers of the Level 2 ARD that corresponds to the given wavelength domain.
The matching table is used to subset / match the spectral bands according to multiple sensors given in the sensor white list.

SENSOR	**Blue**	**Green**	**Red**	**Red Edge 1**	**Red Edge 2**	**Red Edge 3**	**Broad NIR**	**Near Infrared**	**Shortwave Infrared 1**	**Shortwave Infrared 2**
Short	BLUE	GREEN	RED	RE1	RE2	RE3	BNIR	NIR	SWIR1	SWIR2
LND04	1	2	3	-	-	-	-	4	5	6
LND05	1	2	3	-	-	-	-	4	5	6
LND07	1	2	3	-	-	-	-	4	5	6
LND08	1	2	3	-	-	-	-	4	5	6
SEN2A	1	2	3	4	5	6	7	8	9	10
SEN2B	1	2	3	4	5	6	7	8	9	10
sen2a	1	2	3	-	-	-	7	-	-	-
sen2b	1	2	3	-	-	-	7	-	-	-

3) QAI screening

SCREEN_QAI

This parameter gives you full control of quality screening. All provided quality flags (see Table 6) can be used individually. Use this option rigorously! The parameter expects a list with all quality flags that should be masked out. Following tags are available:

NODATA, CLOUD_OPAQUE, CLOUD_BUFFER, CLOUD_CIRRUS, CLOUD_SHADOW, SNOW, WATER, AOD_FILL, AOD_HIGH, AOD_INT, SUBZERO, SATURATION, SUN_LOW, ILLUMIN_NONE, ILLUMIN_POOR, ILLUMIN_LOW, SLOPED, WVP_NONE.

For details about the quality flags, see section VI.B.5).

4) Spatial properties

X_TILE_MIN, Y_TILE_MIN, X_TILE_MAX, Y_TILE_MAX

These parameters indicate the composite extent and represent the western, northern, eastern and southern boundary of the requested composite. The parameters are interpreted as integer values.

DO_TILE

This parameter controls whether the Level 3 data should be output in the gridded data structure used by FORCE L2PS (TRUE) or if one large mosaic of composites/statistics should be generated (FALSE). Due to performance considerations, it is advised to use tiled output; see also *force-level3-mosaic* for generating mosaics after running *force-level3*.

RESOLUTION

This parameter gives the spatial resolution of the resulting Level 3 products (in projection units, commonly in meters), which is e.g. necessary if multi-sensor composited are to be generated. Image decimation/replication is taken care of using nearest neighbor resampling, unless REDUCE_PSF = TRUE (see next parameter). RESOLUTION is interpreted as float value.

REDUCE_PSF

This parameter is related to RESOLUTION, and indicates whether image decimation (i.e. spatial degradation) is done using approximated Point Spread Functions (PSF) – instead of the default nearest neighbor resampling. The PSFs are parameterized with Full Width at Half Maximum (FWHM) corresponding to the size of one pixel at the target resolution. Although an approximation, results are much more realistic than the default behavior. Note that time and processing required for input is considerable, especially if spatial resolution is degraded much.

USE_IMPROPHE

This parameter indicates whether the analysis should be performed on the original BOA reflectance (FALSE) – or on the product with improved spatial resolution (TRUE). If USE_IMPROPHE = TRUE, the IMP product is read instead of the BOA product. If no IMP product is available for any date, the BOA product is read. For improving the spatial resolution of lower resolution Level 2 ARD using higher resolution Level 2 ARD, refert to section VI.G.

5) Inter-annual suitability

YEAR_TARGET

This parameter specifies the target year for compositing. This parameter is interpreted as integer value.

YEAR_NUM

This parameter specifies the number of bracketing years (target year ± bracketing years), i.e. the compositing period. A value of 2 would result in a five-year compositing period. This parameter is interpreted as integer value.

Y_FACTOR

This parameter is a tradeoff parameter that balances the inter- and intra-annual selection. Lower values (e.g. 0.75) favor data from the target year. Higher values favor data that was acquired close to the target DOY (regardless of the year). This parameter is interpreted as float value.

6) Intra-annual suitability

`DOY_SCORE_0, DOY_SCORE_1, DOY_SCORE_2`

These parameters specify the function values $s_{0–2}$ used for fitting the DOY scoring functions. The function type is automatically chosen from the given values:

Gaussian	$s_0 < s_1 > s_2$
Descending sigmoid	$s_0 > s_1 > s_2$
Ascending sigmoid	$s_0 < s_1 < s_2$

Due to numerical reasons, values should be in the range 0.01–1.0 for the Gaussian, and 0.01–0.99 for the sigmoids.

`DOY_STATIC_0, DOY_STATIC_1, DOY_STATIC_2`

These parameters specify the DOYs $p_{0–2}$ used for fitting the DOY scoring functions in case of the static compositing. They are not used for the phenology-adaptive compositing.

However, in each case, the target date appearing in the file name is derived from these values. The target date is the $p_{0–2}$ with highest $s_{0–2}$ (see last parameter).

The DOYs should be in the range 1–366. Typically, $p_0 < p_1 < p_2$, e.g. $p_0 = 60$, $p_1 = 90$, $p_2 = 120$. However, the DOY scoring can also extend between the years (i.e. around the turn of the year). If $p_0 > p_1$: p_0 is from previous year, e.g. $p_0 = 330$, $p_1 = 30$, $p_2 = 90$. If $p_2 < p_1$: p_2 is from next year, e.g. $p_0 = 300$, $p_1 = 330$, $p_2 = 30$.

`OFF_SEASON`

This parameter specifies whether all available data from the requested years are used – or only from the season of interest. This parameter has important implications for both composites and spectral temporal metrics.

If FALSE, the composites only consider data for the period, in which the intra-annual score is higher than 0.01. If there is no clear-sky data within this period, data gaps are possible. The spectral temporal metrics are only calculated for this season as well. This improves the seasonal integrity of the composites and enables the generation of spectral-temporal metrics per season. However, the number of available observations is reduced, which may affect composites (gaps) and spectral temporal metrics that may appear noisier due to a lower number of observations.

If TRUE, all data from the requested years are used, thus the risk of having data gaps is lower and the spectral temporal metrics are computed from more data, which may reduce noise. However, it is possible that data from unwanted parts of the year are selected. The spectral temporal metrics are calculated for the entire year(s). Thus, the same metrics will be calculated for different target dates, which discards seasonal information. Be aware that a different data availability in different seasons between years may affect the inter-annual comparability of these metrics.

7) Phenology-adaptive compositing

`LSP_DO`

This parameter indicates whether the phenology-adpative compositing (TRUE) or the static compositing (FALSE) should be used. In case of the static version, the target DOYs are derived from DOY_STATIC_0, DOY_STATIC_1 and DOY_STATIC_2; and the remaining parameters in this subsection are ignored. In the case of the phenology-adaptive version, a Land Surface Phenology dataset must be prepared (see section VI.C.2), specified with DIR_LSP; and the parameters in this subsection must be specified.

`LSP_PATTERN_PAR_0, LSP_PATTERN_PAR_1, LSP_PATTERN_PAR_2`

These parameters specify the LSP datasets that are used as compositing targets (analogously to DOY_STATIC_0, DOY_STATIC_1, DOY_STATIC_2). One file needs to be prepared for each seasonal parameter (see sectionVI.C.2), and the filenames need to contain a unique ID, which needs to be specified in the parameter file. Example:

`LSP_PATTERN_PAR_0 = POS`	filename: LSP-SEN2A-POS.tif
`LSP_PATTERN_PAR_1 = EOS`	filename: LSP-SEN2A-EOS.tif
`LSP_PATTERN_PAR_2 = MOS`	filename: LSP-SEN2A-MOS.tif

`LSP_1ST_YEAR`

This parameter is the year, which corresponds to the 1^{st} band in the LSP dataset.

LSP_START

This parameter specifies the starting point of the LSP values. Internally, the data are represented as '*Year · 365 + DOY*'. Thus, LSP_START is an offset, which must be given as '*Year · 365 + DOY*'. If the values are provided in this format, use LSP_START = 1. If the LSP values would be provided relative to January 1 2000, use LSP_START = 730001, i.e. 2000*365+1. Leap years are not taken into account and each year consists of 365 days.

LSP_THRESHOLD

This parameter is a threshold in days. If the inter-annual variability of the LSP (of a given pixel) exceeds this value, the long-term average LSP is used instead of the yearly values. The value should be between 0 (long-term average is used for all pixels) and 365 (long-term average is never used).

LSP_NODATA

This parameter specifies the nodata value of the LSP dataset.

8) Compositing scores

SCORE_DOY_WEIGHT

This parameter controls the strength of the DOY score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

SCORE_YEAR_WEIGHT

This parameter controls the strength of the Year score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

SCORE_CLOUD_WEIGHT

This parameter controls the strength of the cloud distance score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

SCORE_HAZE_WEIGHT

This parameter controls the strength of the haze score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

SCORE_CORREL_WEIGHT

This parameter controls the strength of the correlation score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

SCORE_VZEN_WEIGHT

This parameter controls the strength of the view zenith score. SCORE_DOY_WEIGHT = 0 disables the use of this score. This parameter is interpreted as float value.

DREQ

This parameter indicates the distance (to the next cloud or cloud shadow) after which the sky is assumed to be clear (cloud score approaches 1.0). The distance needs to be given in meters. This parameter is interpreted as integer value.

VREQ

This parameter indicates the view zenith angle at which the view zenith score approaches 0.0. The angle needs to be given in degree. This parameter is interpreted as float value.

9) Miscellaneous

NUM_CPU

This parameter controls how many CPUs should be used for parallel processing. This parameter is interpreted as integer value.

10) Output options

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

`OUTPUT_BAP`

This parameter controls whether the best available pixel product should be output. Valid values are TRUE / FALSE.

`OUTPUT_INF`

This parameter controls whether the compositing information product should be output. Valid values are TRUE / FALSE.

`OUTPUT_SCR`

This parameter controls whether the compositing score product should be output. Valid values are TRUE / FALSE.

`OUTPUT_AVG`

This parameter controls whether the temporal average product should be output. Valid values are TRUE / FALSE.

`OUTPUT_STD`

This parameter controls whether the temporal standard deviation product should be output. Valid values are TRUE / FALSE.

`OUTPUT_MIN`

This parameter controls whether the temporal minimum product should be output. Valid values are TRUE / FALSE.

`OUTPUT_MAX`

This parameter controls whether the temporal maximum product should be output. Valid values are TRUE / FALSE.

`OUTPUT_RNG`

This parameter controls whether the temporal range product should be output. Valid values are TRUE / FALSE.

`OUTPUT_SKW`

This parameter controls whether the temporal skewness product should be output. Valid values are TRUE / FALSE.

`OUTPUT_KRT`

This parameter controls whether the temporal kurtosis product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q25`

This parameter controls whether the temporal 0.25 quantile product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q50`

This parameter controls whether the temporal 0.50 quantile product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q75`

This parameter controls whether the temporal 0.75 quantile product should be output. Valid values are TRUE / FALSE.

`OUTPUT_IQR`

This parameter controls whether the temporal interquartile range product should be output. Valid values are TRUE / FALSE.

E. TSA parameter file

The TSA parameter file is mandatory for FORCE TSA. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_TSA_START++ and ++PARAM_TSA_END++ keywords enclose the parameter file. All parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-tsa*.

1) Directories

`DIR_LEVEL2`

This is the input directory, where the Level 2 ARD are stored.

`DIR_MASK`

This is the directory, where analysis masks are stored. Analysis masks are optional and may be disabled by setting the parameter to NULL.

`DIR_TSA`

This is the output directory, where the time series products will be stored.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be processed/output. If FILE_TILE = NULL, all available tiles (within the spatial extent; see below) are output.

`FILE_ENDMEM`

This is the endmember file for Spectral Mixture Analysis (see section VII.F). If INDEX != SMA, FILE_ENDMEM may be NULL.

2) Sensor white list

`SENSORS`

This parameter expects a list of sensor IDs that should be used for generating / analyzing the time series. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = LND07 LND08 SEN2A SEN2B
```

You will be notified if the selected index is not available for the chosen sensor set. Note that no spectral adjustment is made. Following sensors are available:

LND04 (6-band Landsat 4 TM), LND05 (6-band Landsat 5 TM), LND07 (6-band Landsat 7 ETM+), LND08 (6-band Landsat 8 OLI), SEN2A (10-band Sentinel-2A), SEN2B (10-band Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). The matching table (Table 8, p. 65) is applied to determine overlapping bands.

3) QAI screening

`SCREEN_QAI`

This parameter gives you full control of quality screening. All provided quality flags (see Table 6) can be used individually. Use this option rigorously! The parameter expects a list with all quality flags that should be masked out. Following tags are available:

NODATA, CLOUD_OPAQUE, CLOUD_BUFFER, CLOUD_CIRRUS, CLOUD_SHADOW, SNOW, WATER, AOD_FILL, AOD_HIGH, AOD_INT, SUBZERO, SATURATION, SUN_LOW, ILLUMIN_NONE, ILLUMIN_POOR, ILLUMIN_LOW, SLOPED, WVP_NONE.

For details about the quality flags, see section VI.B.5).

4) Spatial properties

`X_TILE_MIN, Y_TILE_MIN, X_TILE_MAX, Y_TILE_MAX`

These parameters indicate the processing extent and represent the western, northern, eastern and southern boundary of the requested area. The parameters are interpreted as integer values.

`RESOLUTION`

This parameter gives the spatial resolution of the resulting time series products (in projection units, commonly in meters), which is e.g. necessary if multi-sensor analyses are to be performed. Image decimation/replication is taken care of using nearest neighbor resampling, unless REDUCE_PSF = TRUE (see next parameter). RESOLUTION is interpreted as float value.

`REDUCE_PSF`

This parameter is related to RESOLUTION, and indicates whether image decimation (i.e. spatial degradation) is done using approximated Point Spread Functions (PSF) – instead of the default nearest neighbor resampling. The PSFs are parameterized with Full Width at Half Maximum (FWHM) corresponding to the size of one pixel at the target resolution. Although an approximation, results are much more realistic than the default behavior. Note that time and processing required for input is considerable, especially if spatial resolution is degraded much.

`USE_IMPROPHE`

This parameter indicates whether the analysis should be performed on the original BOA reflectance (FALSE) – or on the product with improved spatial resolution (TRUE). If USE_IMPROPHE = TRUE, the IMP product is read instead of the BOA product. If no IMP product is available for any date, the BOA product is read. For improving the spatial resolution of lower resolution Level 2 ARD using higher resolution Level 2 ARD, refert to section VI.G.

5) Temporal properties

`DOY_MIN, DOY_MAX`

These parameters give the range of DOYs that should be used for time series generation / analysis. This range is interpreted together with the MONTH_MIN/MONTH_MAX, and the most restrictive seasonal subset is used. As an example, if the min–max days are set to 1–200 and the min–max month are set to 3–12, only acquisitions between DOY 60 and 200 are used. The parameters are interpreted as integer values.

`MONTH_MIN, MONTH_MAX`

These parameters give the range of months that should be used for time series generation / analysis. This range is interpreted together with the DOY_MIN/DOY_MAX, and the most restrictive seasonal subset is used. As an example, if the min–max days are set to 1–200 and the min–max month are set to 3–12, only acquisitions between DOY 60 and 200 are used. The parameters are interpreted as integer values.

`YEAR_MIN, YEAR_MAX`

These parameters give the range of years that should be used for time series generation / analysis. The parameters are interpreted as integer values.

6) Time series parameters

`INDEX`

This parameter specifies which spectral index should be used to build the time series. You will be notified if the selected index is not available for the chosen sensor set.

Valid keys are BLUE (blue band), GREEN (green band), RED (red band), RE1 (Red Edge band 1), RE2 (Red Edge band 2), RE3 (Red Edge band 3), BNIR (broad near infrared band), NIR (near infrared band), SWIR1 (shortwave infrared band 1), SWIR2 (shortwave infrared band 2), NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), NBR (Normalized Burn Ratio), ARVI (Atmospherically Resistant Vegetation Index), SAVI (Soil Adjusted Vegetation Index), SARVI (Soil and Atmospherically Resistant Vegetation Index), TC-BRIGHT (Tasseled Cap Brightness), TC-GREEN (Tasseled Cap Greenness), TC-WET (Tasseled Cap Wetness), TC-DI (Tasseled Cap-based Disturbance Index), TC-DI0 (Tasseled Cap-based Disturbance Index without rescaling), SMA (Spectral Mixture Analysis).

CENTER

This parameter indicates whether the time series are centered on the pixel means before output (see Fig. 5). Valid values are TRUE / FALSE.

STANDARD

This parameter indicates whether the time series are standardized with the pixel means and standard deviations before output (see Fig. 5). Valid values are TRUE / FALSE.

TREND

This parameter indicates the aggregation unit for time series folding (aggregation). The time series can be folded by YEAR (annual time series; this is probably the method of choice for most applications), MONTH, WEEK or DAY.

FOLD

This parameter indicates the folding method used to aggregate the time series by year, month, week or day. Valid keys are AVG (average), MIN (minimum), MAX (maximum), or LSP (Land Surface Phenology).

TAIL

This parameter indicates the tail type used for testing whether the slope of the trend analysis is significantly different from zero (t-test). Valid keys are TWO, LEFT, RIGHT.

CONF

This parameter indicates the confidence level (e.g. 0.95) used for testing whether the slope of the trend analysis is significantly different from zero. This parameter is interpreted as float value.

DESPIKE

This option can be used to remove spikes (i.e. positive outliers) from the time series. Each datapoint is predicted with linear interpolation using the two bracketing datapoints. A spike is detected if the deviation between actual data and predicted data is larger than the value given by DESPIKE. This parameter is interpreted as float value and needs to be given in unscaled index units (e.g. INDEX = SMA & DESPIKE = 0.1 means that any datapoint, whose fraction is 10% larger than the bracketing datapoints, will be masked out).

7) Spectral unmixing properties

SMA_SUM_TO_ONE

This parameter indicates whether the SMA should be Sum-to-One constrained. Valid values are TRUE / FALSE.

SMA_NON_NEG

This parameter indicates whether the SMA should be non-negativity constrained using the fast non-negativity-constrained least squares algorithm. This option has an effect on processing time. Valid values are TRUE / FALSE.

SMA_SHD_NORM

This parameter indicates whether the SMA fractions should be shade-normalized. If so, a shade endmember has to be given in the last column of the endmember file (see VII.F).You may want to filter water pixels (see SCREEN_QAI). Valid values are TRUE / FALSE.

SMA_ENDMEMBER

Only the abundance image related to one endmember is retained for output / further processing. This parameter indicates, which endmember to use (i.e. the column in the endmember file, starting with 1, see VII.F). This parameter is interpreted as integer value.

8) Interpolation properties

INTERPOLATE

This parameter specifies the interpolation method. Currently available are no interpolation (NONE) linear interpolation (LINEAR), moving average filter (MOVING), and Radial Basis Function (RBF) ensembles.

Note that all interpolation methods consider time. Depending on data density, the method, and on the following settings, nodata values are still possible after interpolation.

If Land Surface Phenology metrics should be output (OUTPUT_LSP = TRUE), or Trend or CAT should be computed based on LSP (FOLD = LSP), INTERPOLATE cannot be NULL.

`MOVING_MAX`

This parameter indicates the maximum temporal distance (in days), which is considered for the moving average. If there is no observation within this distance, a nodata value will be present in the interpolated time series. This parameter is interpreted as integer value.

`RBF_SIGMA`

This parameter expects a list of kernel widths for the Radial Basis Function ensembles. The kernel widths are the sigmas (σ) of the Gaussian functions in days, e.g. SIGMA = 8 16 32. The RBF kernel strengths are adapted by weighting with actual data availability within the time period covered by each kernel. RBF ensembles are useful for time series that include periods with fairly sparse and dense data availability. This parameter is interpreted as a list of integer values.

`RBF_CUTOFF`

This parameter indicates the cutoff density for the RBF kernels. As an example, RBF_CUTOFF = 0.95 will truncate the RBF kernels, such that each kernel is long enough to include 95% of the area under its Gaussian. The actual time depends on RBF_SIGMA. The maximum value is 1, i.e. the entire time series is considered to predict each datapoint (not recommended, this will be pretty slow…). If there is no observation within the truncated Gaussian of the broadest RBF kernel, a nodata value will be present in the interpolated time series. This parameter is interpreted as float value.

`INT_DAY`

This parameter indicates the prediction step in days, i.e. the temporal resolution of the interpolated time series. This parameter is interpreted as integer value.

9) Phenology properties

`LSP_DOY_PREV_YEAR, LSP_DOY_NEXT_YEAR`

To derive LSP metrics for a given year, some data from the previous and next year need to be included, as the early / late minima may not be found in the reference year (the year for which LSP metrics should be extracted). The minima are needed to derive all the LSP metrics in between. These parameters specify the range of previous / next year data that should be considered. Values are in DOY. As an example for the reference year 2015, LSP_DOY_PREV_YEAR = 91 and LSP_DOY_NEXT_YEAR = 182 would consider all data from April 2014 to June 2016. These parameters are interpreted as integer values.

`LSP_HEMISPHERE`

This parameter indicates if you are expecting a Northern hemisphere phenology (0: peak of season is in the middle of the year), a Southern hemisphere phenology (1: peak of season is around the turn of the year. For a given reference year, the corresponding POS shall be at the end of the previous year or at the beginning of the reference year), or a mixed type where both can happen (2). In the latter case, FORCE TSA analyzes the time series of each pixel, and then switches to Northern or Southern hemisphere retrieval. This parameter is interpreted as integer value.

`LSP_N_SEGMENT`

This parameter specifies the number of spline segments per 365 days. More segments give a more detailed fit; the detail vs. smoothing capability of the spline needs to be considered with respect to application, land cover of interest and data density. This parameter is interpreted as integer value.

10) Miscellaneous options

`NUM_CPU`

This parameter controls how many CPUs should be used for parallel processing. This parameter is interpreted as integer value.

11) Output options

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

`OUTPUT_TSS`

This parameter controls whether the time series stack product should be output. Valid values are TRUE / FALSE.

`OUTPUT_RMS`

This parameter controls whether the RMSE time series of the SMA should be output. Valid values are TRUE / FALSE.

`OUTPUT_STA`

This parameter controls whether the basic statistics product should be output. Valid values are TRUE / FALSE.

`OUTPUT_TSI`

This parameter controls whether the time series interpolation product should be output. Valid values are TRUE / FALSE.

`OUTPUT_FBY`

This parameter controls whether the fold-by-year time series stack product should be output. Valid values are TRUE / FALSE.

`OUTPUT_FBM`

This parameter controls whether the fold-by-month time series stack product should be output. Valid values are TRUE / FALSE.

`OUTPUT_FBW`

This parameter controls whether the fold-by-week time series stack product should be output. Valid values are TRUE / FALSE.

`OUTPUT_FBD`

This parameter controls whether the fold-by-day time series stack product should be output. Valid values are TRUE / FALSE.

`OUTPUT_TRD`

This parameter controls whether the trend analysis product should be output. Valid values are TRUE / FALSE.

`OUTPUT_CAT`

This parameter controls whether the change and trend analysis product should be output. Valid values are TRUE / FALSE.

`OUTPUT_LSP`

This parameter controls whether the 26 Land Surface Phenology products should be output. Valid values are TRUE / FALSE.

F. Endmember file

This file is needed for Spectral Mixture Analysis in FORCE TSA, i.e. if INDEX = SMA. Suggested file extension is '.emb'. The file defines the endmember spectra that should be used for the SMA. The values need to be given in scaled reflectance (scale factor 10,000). The files should be without header, ended with an empty line, columns separated by white-space.

There should be one column for each endmember. If you want to apply shade normalization, the shade spectrum (photogrammetric zero or measured shade) needs to be in the last column. There should be as many lines as there are overlapping bands for the chosen set of sensors. See Table 8 on p. 65 for the matching table.

As an example, generating a fraction time series based on LND04, LND05, LND07, and LND08 requires 6-band endmembers (Landsat legacy bands). Generating a fraction time series based on LND08, SEN2A and SEN2B requires 6-band endmembers (Landsat legacy bands). Generating a fraction time series based on SEN2A and SEN2B requires 10-band endmembers (Sentinel-2 land surface bands). Generating a fraction time series based on sen2a and sen2b requires 4-band endmembers (Sentinel-2 high-res bands).

Example (Landsat legacy bands using vegetation, soil, rock and shade endmembers):

```
320 730 2620 0
560 1450 3100 0
450 2240 3340 0
3670 2750 4700 0
1700 4020 7240 0
710 3220 5490 0
```

G. CSO parameter file

The CSO parameter file is mandatory for FORCE CSO. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_CSO_START++ and ++PARAM_CSO_END++ keywords enclose the parameter file. All parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-cso*.

1) Directories

`DIR_LEVEL2`

This is the input directory, where the Level 2 ARD are stored.

`DIR_CSO`

This is the output directory, where the CSO products will be stored.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be processed/output. If FILE_TILE = NULL, all available tiles (within the spatial extent; see below) are output.

2) Sensor white list

`SENSORS`

This parameter expects a list of sensor IDs that should be used for extracting the clear sky observation statistics. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = LND07 LND08 SEN2A SEN2B
```

Following sensors are available:

LND04 (Landsat 4 TM), LND05 (Landsat 5 TM), LND07 (Landsat 7 ETM+), LND08 (Landsat 8 OLI), SEN2A (Sentinel-2A), SEN2B (Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). Only the QAI product is analyzed.

3) QAI screening

`SCREEN_QAI`

This parameter gives you full control of quality screening. All provided quality flags (see Table 6) can be used individually. Use this option rigorously! The parameter expects a list with all quality flags that should be masked out. Following tags are available:

NODATA, CLOUD_OPAQUE, CLOUD_BUFFER, CLOUD_CIRRUS, CLOUD_SHADOW, SNOW, WATER, AOD_FILL, AOD_HIGH, AOD_INT, SUBZERO, SATURATION, SUN_LOW, ILLUMIN_NONE, ILLUMIN_POOR, ILLUMIN_LOW, SLOPED, WVP_NONE.

For details about the quality flags, see section VI.B.5).

4) Spatial properties

`X_TILE_MIN, Y_TILE_MIN, X_TILE_MAX, Y_TILE_MAX`

These parameters indicate the processing extent and represent the western, northern, eastern and southern boundary of the requested area. The parameters are interpreted as integer values.

`RESOLUTION`

This parameter gives the spatial resolution of the resulting CSO products (in projection units, commonly in meters), which is e.g. necessary if multi-sensor analyses are to be performed. Image decimation/replication is taken care of using nearest neighbor resampling. RESOLUTION is interpreted as float value.

5) Temporal properties

`YEAR_MIN, YEAR_MAX`

These parameters give the range of years that should be analyzed. The parameters are interpreted as integer values.

`MONTH_STEP`

This parameter gives the step width for the analysis. CSO statistics are generated for each of these bins. As an example, if YEAR_MIN = 2014, YEAR_MAX = 2016, and MONTH_STEP = 3, CSO statistics will be generated quarterly over a three-year period, resulting in 12 bins. This parameter is interpreted as integer value.

6) Miscellaneous options

`NUM_CPU`

This parameter controls how many CPUs should be used for parallel processing. This parameter is interpreted as integer value.

7) Output options

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

`OUTPUT_NUM`

This parameter controls whether the number of observations product should be output. Valid values are TRUE / FALSE.

`OUTPUT_AVG`

This parameter controls whether the average of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_STD`

This parameter controls whether the standard of dt deviation product should be output. Valid values are TRUE / FALSE.

`OUTPUT_MIN`

This parameter controls whether the minimum of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_MAX`

This parameter controls whether the maximum of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_RNG`

This parameter controls whether the range of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_SKW`

This parameter controls whether the skewness of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_KRT`

This parameter controls whether the kurtosis of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q25`

This parameter controls whether the 0.25 quantile of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q50`

This parameter controls whether the 0.50 quantile of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_Q75`

This parameter controls whether the 0.75 quantile of dt product should be output. Valid values are TRUE / FALSE.

`OUTPUT_IQR`

This parameter controls whether the interquartile range of dt product should be output. Valid values are TRUE / FALSE.

H. ImproPhe parameter file

The ImproPhe parameter file is mandatory for FORCE ImproPhe. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_IMP_START++ and ++PARAM_IMP_END++ keywords enclose the parameter file. All parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-improphe*.

1) Directories and file handling

`DIR_LEVEL2`

This is the input directory, where the Level 2 ARD are stored.

`FILE_COARSE`

This parameter expects a list with all file names (full file path) of the coarse continuous fields that should be ImproPhed. Each product needs to be a separate file. The files can be multiband images, and the bands need to represent a year.

As an example with two continuous fields, the 'Value of the Peak of Season' phenometric, and the 'Integral of the Green Season' phenometric should be given as separate multiband files. Each band represents the LSP of one year, e.g. from 2000–2015.

The files should cover the complete study area. There is not really a restriction on the projection or resolution; the file format must be supported by *GDAL*. The files will be warped to the extent and resolution of the processed image using nearest neighbor resampling.

`PRODUCT_TAG`

This parameter expects a list with 3-digit product tags, which are used to name the output images, see VI.F.3).

`DIR_IMPROPHE`

This is the output directory, where the ImproPhed continuous fields will be stored.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be processed/output. If FILE_TILE = NULL, all available tiles (within the spatial extent; see below) are output.

2) Sensor white list

`SENSORS`

This parameter expects a list of sensor IDs that should be used for ImproPhing the coarse continuous fields. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = LND07 LND08 SEN2A SEN2B
```

Note that no spectral adjustment is made. Following sensors are available:

LND04 (6-band Landsat 4 TM), LND05 (6-band Landsat 5 TM), LND07 (6-band Landsat 7 ETM+), LND08 (6-band Landsat 8 OLI), SEN2A (10-band Sentinel-2A), SEN2B (10-band Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). The matching table (Table 8, p. 65) is applied to determine overlapping bands.

3) QAI screening

`SCREEN_QAI`

This parameter gives you full control of quality screening. All provided quality flags (see Table 6) can be used individually. Use this option rigorously! The parameter expects a list with all quality flags that should be masked out. Following tags are available:

NODATA, CLOUD_OPAQUE, CLOUD_BUFFER, CLOUD_CIRRUS, CLOUD_SHADOW, SNOW, WATER, AOD_FILL, AOD_HIGH, AOD_INT, SUBZERO, SATURATION, SUN_LOW, ILLUMIN_NONE, ILLUMIN_POOR, ILLUMIN_LOW, SLOPED, WVP_NONE.

For details about the quality flags, see section VI.B.5).

4) Spatial properties

`X_TILE_MIN, Y_TILE_MIN, X_TILE_MAX, Y_TILE_MAX`

These parameters indicate the processing extent and represent the western, northern, eastern and southern boundary of the requested area. The parameters are interpreted as integer values.

`RESOLUTION`

This parameter gives the spatial resolution of the resulting continuous fields (in projection units, commonly in meters). Image decimation/replication is taken care of using nearest neighbor resampling. RESOLUTION is interpreted as float value.

`REDUCE_PSF`

This parameter is related to RESOLUTION, and indicates whether image decimation (i.e. spatial degradation) is done using approximated Point Spread Functions (PSF) – instead of the default nearest neighbor resampling. The PSFs are parameterized with Full Width at Half Maximum (FWHM) corresponding to the size of one pixel at the target resolution. Although an approximation, results are much more realistic than the default behavior. Note that time and processing required for input is considerable, especially if spatial resolution is degraded much.

5) Temporal properties

`YEAR_MIN, YEAR_MAX`

These parameters give the range of years that should be ImproPhed. The parameters are interpreted as integer values.

`SEASONAL_WINDOW`

This parameter defines the seasonal windows for which the medium resolution data should be aggregated. This parameter expects a list of DOYs that define the window breakpoints. If you specify 5 breakpoints, there will be four windows. The parameters are interpreted as integer values.

There should be enough windows to capture the general phenological differences between land covers (e.g. four windows). The windows should be broad enough to have a high probability of valid data. It is not necessary to use very dense windows (if you have enough data for this, you might want to derive LSP directly from this data).

The windows can be defined in the previous year, in which case a negative number needs to be given; e.g. -15 is a breakpoint in the middle of December in the previous year. The windows can be defined in the next year, in which case a number larger than 365 needs to be given; e.g. 380 is a breakpoint in the middle of January in the next year.

6) Prediction options

`KERNEL_SIZE`

This parameter indicates the width (i.e. diameter) of the prediction kernel (in projection units, commonly in meters). This parameter is interpreted as float value.

`KERNEL_TEXT`

This parameter indicates the width (i.e. diameter) of the kernel used for computing the heterogeneity proxies (in projection units, commonly in meters). The heterogeneity proxies are derived from a focal standard deviation filter. The width of the kernel should reflect the scale difference between the coarse and medium resolution data. This parameter is interpreted as float value.

7) Coarse resolution options

`COARSE_1ST_YEAR`

This parameter is the year, which corresponds to the 1st band in the coarse continuous field dataset(s). This parameter is interpreted as integer value.

`COARSE_NODATA`

This parameter specifies the nodata value of the coarse continuous field dataset(s). This parameter is interpreted as float value.

8) Miscellaneous options

`NUM_CPU`

This parameter controls how many CPUs should be used for parallel processing. This parameter is interpreted as integer value.

9) Output options

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

I. L2IMP parameter file

The Level 2 ImproPhe parameter file is mandatory for FORCE L2IMP. The file extension is '.prm'. All parameters must be given, even if they are not used. Rudimentary checks are performed by the software components using this file. The ++PARAM_L2IMP_START++ and ++PARAM_L2IMP_END++ keywords enclose the parameter file. All parameters follow common tag-value notation with float, integer or character data types, separated by ' = '. The following subsections briefly describe the available parameters. A parameter file skeleton can be generated with *force-parameter-l2imp*.

1) Directories

`DIR_LEVEL2`

This is the input / output directory, where the Level 2 ARD are stored, and where the ImproPhed products will be stored.

`FILE_TILE`

This tile white list (see section VII.C) specifies tiles that are to be processed/output. If FILE_TILE = NULL, all available tiles (within the spatial extent; see below) are output.

2) Sensor white list

`SENSORS_MR`

This parameter expects a list of medium resolution sensor IDs. The spatial resolution of the BOA images corresponding to these sensors will be improved. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = LND07 LND08
```

Note that no spectral adjustment is made. Following sensors are available.

LND04 (6-band Landsat 4 TM), LND05 (6-band Landsat 5 TM), LND07 (6-band Landsat 7 ETM+), LND08 (6-band Landsat 8 OLI), SEN2A (10-band Sentinel-2A), SEN2B (10-band Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). The matching table (Table 8, p. 65) is applied to determine overlapping bands.

`SENSORS_HR`

This parameter expects a list of high resolution sensor IDs. The BOA images corresponding to these sensors will be used as targets for the ImproPhement. All sensors need to be given in a single line, separated by white space. Example with current satellite data:

```
SENSORS = sen2a sen2b
```

Note that no spectral adjustment is made. Following sensors are available.

LND04 (6-band Landsat 4 TM), LND05 (6-band Landsat 5 TM), LND07 (6-band Landsat 7 ETM+), LND08 (6-band Landsat 8 OLI), SEN2A (10-band Sentinel-2A), SEN2B (10-band Sentinel-2B), sen2a (4-band Sentinel-2A), sen2b (4-band Sentinel-2B). The matching table (Table 8, p. 65) is applied to determine overlapping bands.

If you want to improve the spatial resolution of Landsat imagery with Sentinel-2 targets, it is advised to use the high-res bands as targets for the data fusion, i.e. SENSORS = sen2a sen2b will give better results than SENSORS = SEN2A SEN2B.

3) QAI screening

`SCREEN_QAI`

This parameter gives you full control of quality screening. All provided quality flags (see Table 6) can be used individually. Use this option rigorously! The parameter expects a list with all quality flags that should be masked out. Following tags are available:

NODATA, CLOUD_OPAQUE, CLOUD_BUFFER, CLOUD_CIRRUS, CLOUD_SHADOW, SNOW, WATER, AOD_FILL, AOD_HIGH, AOD_INT, SUBZERO, SATURATION, SUN_LOW, ILLUMIN_NONE, ILLUMIN_POOR, ILLUMIN_LOW, SLOPED, WVP_NONE.

For details about the quality flags, see section VI.B.5).

4) Spatial properties

`X_TILE_MIN, Y_TILE_MIN, X_TILE_MAX, Y_TILE_MAX`

These parameters indicate the processing extent and represent the western, northern, eastern and southern boundary of the requested area. The parameters are interpreted as integer values.

`RESOLUTION`

This parameter gives the spatial resolution at which the ImproPhement should be performed (in projection units, commonly in meters). If you want to improve the spatial resolution of Landsat imagery with Sentinel-2 targets, this parameter is usually 10. Image decimation/replication is taken care of using nearest neighbor resampling. RESOLUTION is interpreted as float value.

`REDUCE_PSF`

This parameter is related to RESOLUTION, and indicates whether image decimation (i.e. spatial degradation) is done using approximated Point Spread Functions (PSF) – instead of the default nearest neighbor resampling. The PSFs are parameterized with Full Width at Half Maximum (FWHM) corresponding to the size of one pixel at the target resolution. Although an approximation, results are much more realistic than the default behavior. Note that time and processing required for input is considerable, especially if spatial resolution is degraded much.

5) Temporal properties

`YEAR`

This parameter gives the year, in which data should be ImproPhed. This parameter is interpreted as integer value.

`SEASONAL_WINDOW`

This parameter defines the seasonal windows for which the medium resolution data should be aggregated. This parameter expects a list of DOYs that define the window breakpoints. If you specify 5 breakpoints, there will be four windows. The parameters are interpreted as integer values.

There should be enough windows to capture the general phenological differences between land covers (e.g. four windows). The windows should be broad enough to have a high probability of valid data. It is not necessary to use very dense windows.

The windows can be defined in the previous year, in which case a negative number needs to be given; e.g. -15 is a breakpoint in the middle of December in the previous year. The windows can be defined in the next year, in which case a number larger than 365 needs to be given; e.g. 380 is a breakpoint in the middle of January in the next year.

6) Prediction options

`KERNEL_SIZE`

This parameter indicates the width (i.e. diameter) of the prediction kernel (in projection units, commonly in meters). This parameter is interpreted as float value.

`KERNEL_TEXT`

This parameter indicates the width (i.e. diameter) of the kernel used for computing the heterogeneity proxies (in projection units, commonly in meters). The heterogeneity proxies are derived from a focal standard deviation filter. The width of the kernel should reflect the scale difference between the coarse and medium resolution data. This parameter is interpreted as float value.

7) Miscellaneous options

`NUM_CPU`

This parameter controls how many CPUs should be used for parallel processing. This parameter is interpreted as integer value.

8) Output options

`OUTPUT_OVERWRITE`

This parameter indicates whether already existing ImproPhements should be overwritten, or if those images are skipped.

`OUTPUT_FORMAT`

This parameter controls whether output is written as flat binary image without compression (ENVI) or if a LZW-compressed GeoTiff with horizontal differencing (GTiff) is written.

J. *Water Vapor look-up-tables*

The usage of a spatially and temporally explicit water vapor database for atmospheric correction in FORCE L2PS is optional for Landsat data. Water vapor LUTs are ignored for Sentinel-2 data, but should be considered for Landsat (especially for the TM-type of sensors). The directory containing the water vapor look-up-tables is specified with the DIR_WVPLUT in the Level 2 parameter file (see section VII.A.1). However, the algorithm can also be parameterized with a global water vapor value, see WATER_VAPOR.

FORCE WVDB provides a software component to generate such a database on the basis of MODIS MOD05/MYD05 data; see section VI.H. Note that the initial build of the water vapor database may need some time. However, the user may also use other data sources and/or tools to generate the required tables. Or you can download an application-ready database from https://doi.org/10.1594/PANGAEA.893109 (global coverage, 2000–July 2018).

There are two sets of tables that can/need to be generated:

1) Daily tables are optional, but recommended. They contain a water vapor value for each coordinate, and there is one table for each day.

2) Climatology tables are mandatory (unless DIR_WVPLUT = NULL). L2PS uses the climatology tables if a daily table is unavailable or if there is a fill value in the daily table. Therefore, the minimum requirement is to prepare the 12 climatology tables, one for each month.

1) Daily tables

For each date, one file can be prepared. The file naming is WVP_YYYY-MM-DD.txt; e.g. WVP_2003-08-24.txt. The files are four column tables with no header, separated by white-space. One line per coordinate; ended with an empty line. The coordinate closest to the scene center will be selected, and the corresponding value will be retrieved.

Longitude (1st column) and latitude (2nd column) need to be given as geographic coordinates in decimal degree (negative values for South/West), followed by water vapor (3rd column), and three-digit source (4th column). The fill value is 9999 and TBD for source. The generation of daily tables is not mandatory, but highly recommended. If there is no table for a specific day, or if there is a fill value in the table, the corresponding climatology table is used instead.

Example:

```
17.2642002 -14.4588003 2.448023 MYD
16.9421997 -15.9028997 2.189836 MYD
20.6735001 -13.0142002 9999.000 TBD
20.3544006 -14.4588003 2.427723 MOD
20.0323009 -15.9028997 2.499933 MOD
… [file truncated]
```

2) Climatology tables

12 climatology tables must be prepared, one per month. The file naming is WVP_0000-MM-00.txt; e.g. WVP_0000-06-00.txt. The files are five column tables with no header, separated by white-space. One line per coordinate; ended with an empty line. The coordinate closest to the scene center will be selected, and the corresponding value will be retrieved.

Longitude (1st column) and latitude (2nd column) need to be given as geographic coordinates in decimal degree (negative values for South/West), followed by long-term water vapor average (3rd column), long-term standard deviation of water vapor (4th column) and number of valid observations used for averaging (5th column). The generation of climatology tables is mandatory (unless DIR_WVPLUT = NULL).

Example:

```
96.4300 34.6138 1.205356 0.398807 446
96.0306 33.1801 1.360043 0.399460 447
95.6409 31.7452 1.442830 0.350363 425
95.2598 30.3093 1.642989 0.276430 311
94.8869 28.8723 4.018294 0.812506 149
94.5214 27.4344 6.426344 0.724956 123
… [file truncated]
```

K. AOD look-up-tables

These Look-up-Tables are optional, and may be used to override FORCE L2PS' internal AOD estimation or to provide backup values if the internal AOD estimation failed for any reason. Potential usages are to employ an AOD climatology or a fixed AOD. The directory (containing the LUTs) is specified with DIR_AOD in the Level 2 parameter file (see section VII.A.1).

For each DOY, one file needs to be prepared (you should prepare 366 files). The file naming is AOD_DOY.txt; e.g. AOD_076.txt. The files are five column tables with no header, separated by white-space. One line per coordinate; ended with an empty line. The coordinate closest to the scene center will be selected, and the corresponding AOD will be retrieved.

Longitude (1st column) and latitude (2nd column) need to be given as geographic coordinates in decimal degree (negative values for South/West), followed by three Ångström coefficients (3rd – 5th column; logarithmic formulation, see below). The first coefficient is the turbidity coefficient (a_0), the second coefficient is the Ångström exponent (a_1), and the third coefficient can be used to describe spectral curvature in AOD (a_2). The spectral curvature can be disabled with $a_2 = 0$; in this case the formulation simplifies to the classic Ångström equation. AOD for any given wavelength is retrieved using following equation:

$$\ln \tau_a = a_0 + a_1 \cdot \ln \lambda + a_2 \cdot (\ln \lambda)^2 \quad (1)$$

L. Coordinate file

This file is needed for some tools, e.g. FORCE WVDB. Suggested file extension is '.coo'. The file defines coordinates that should be processed with some functionality. The files are two column tables with no header, separated by white-space. One line per coordinate; ended with an empty line. Longitude (1st column) and latitude (2nd column) need to be given as geographic coordinates in decimal degree (negative values for South/West).

Example:

```
17.2642 -14.4588
16.9422 -15.9029
20.6735 -13.0142
20.3544 -14.4588
20.0323 -15.9029
… [file truncated]
```

M. Data Cube definition

The spatial data cube definition is appended to each data cube, i.e. to each directory containing tiled datasets. The file 'datacube-definition.prj' is a 6-line text file that contains the (1) projection as WKT string, (2) origin of the tile system as geographic Longitude, (3) origin of the tile system as geographic Latitude, (4) origin of the tile system as projected X-coordinate, (5) origin of the tile system as projected Y-coordinate, and (6) width of the tiles in projection units. Do not modify or delete any of these files!

Example (WKT string is one line):

```
PROJCS["ETRS89 / LAEA
Europe",GEOGCS["ETRS89",DATUM["European_Terrestrial_Reference_System_1989",SPHERO
ID["GRS 1980",6378137,298.257222101,AUTHORITY["EPSG","7019"]],TOWGS84[0,0,0,0,0,0
,0],AUTHORITY["EPSG","6258"]],PRIMEM["Greenwich",0,AUTHORITY["EPSG","8901"]],UNIT
["degree",0.0174532925199433,AUTHORITY["EPSG","9122"]],AUTHORITY["EPSG","4258"]],
PROJECTION["Lambert_Azimuthal_Equal_Area"],PARAMETER["latitude_of_center",52],PAR
AMETER["longitude_of_center",10],PARAMETER["false_easting",4321000],PARAMETER["fa
lse_northing",3210000],UNIT["metre",1,AUTHORITY["EPSG","9001"]],AUTHORITY["EPSG",
"3035"]]
-25.000000
60.000000
2456026.250000
4574919.500000
30000.000000
```

VIII. PROJECTION AND TILING

The choice of the output projection is of key importance for generating ARD, but choosing an appropriate one is not trivial and often leads to confusion (especially with additional tiling). This chapter is intended to give some guidance and ideas on how to choose reasonable projection and tiling parameters.

First of all, in most cases, it is good practice to reproject the data to one shared coordinate system because the space agencies ship data in UTM projection with different zones. Thus, if the input data cover different UTM zones, the output data cannot be co-registered easily if reprojection is disabled. Note that tiling is still a valid option, but the results should be used with extreme care. Tiling primarily enables pixel-based operations, but these should not be used with a DO_REPROJ = FALSE / DO_TILE = TRUE parameterization (except for areas that are covered by one UTM zone only).

There exist many projections, and custom projections are also allowed (http://spatialreference.org can be used to find an existing one). We use the *GDAL* library for reprojection purposes. As such, virtually any projection expressed as Well-Known-Text is valid. However, the appropriate projection should be selected with care. In general, the projection should flatten your study area with minimal distortion. The choice depends on the location, size and extent of your study area, as well as on the desired specifications. Frequently, large area production systems use equal area projections with different projection surfaces for different study areas (e.g. Albers Conic → CONUS or Lambert Azimuthal → Pan-European). Snyder's handbook on map projections[2] gives useful recommendations.

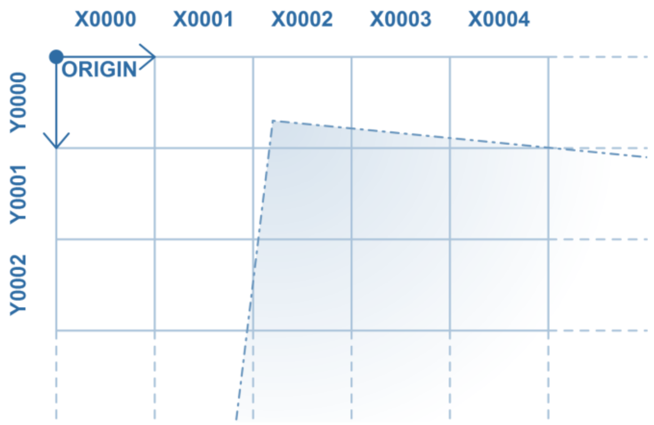

Fig. 12. Level 2 Gridding.
The image is intersected with the arbitrary grid and the chips are extracted. Tile X0000_Y0000, pixel 0/0 starts at ORIGIN_LON/ORIGIN_LAT. The tiles are square, the tile size is given by TILE_SIZE in output map units. TILE_SIZE must be a multiple of RESOLUTION.

After reprojection, the data can be tiled to an arbitrary grid – similar to the MODIS Land products. Pixel-based operations can be easily used if this option is used. In fact, gridding is necessary for all higher-level FORCE operations (> Level 2). The grid can be freely defined – with consideration of the employed projection. We use square tiles (def. tile: grid cell) and the tile size is specified by the TILE_SIZE parameter, which must be given in output projection units (commonly in meters). TILE_SIZE must be a multiple of RESOLUTION.

The tile originates at ORIGIN_LON / ORIGIN_LAT, which must be given in geographic coordinates. These coordinates are transformed to the output projection (or input projection if DO_REPROJ = FALSE). The tile numbers increase from West to East and North to South. A new tile is generated each TILE_SIZE units. Tile X0000_Y0000, pixel 0/0 is located at ORIGIN_LON/ORIGIN_LAT. Negative tile numbers can occur if the tile origin is not North-West of your study area. Although negative tiles are generally not problematic, we recommend to use a tile origin that is sufficiently far away in the North-West of your study area and does not intersect with any input data (but close enough to be represented reasonably in the projected coordinate system).

Note that a geographic location in the North-West is not necessarily North-West in the output coordinate system, too (see Azimuthal example below). This may result in unexpected – yet valid – behavior. Note that coordinates may be undefined if they are too far away from the origin of the coordinate system (see Transverse Mercator example below); in this case the algorithm will fail. It is good practice to use a point of origin that is relatively close to the study area. The allowed distance varies greatly with different output projections.

See also *force-tabulate-grid* for creating a shapefile with the grid (for visualization purposes) and *force-tile-finder* for identifying the tile ID and pixel coordinate of any geographic coordinate.

Finally, the images are intersected with the grid, then defined as chips. The chips are extracted and saved as individual datasets. Empty tiles (e.g. black image boundary) will not be saved. Cloudy tiles can also be suppressed (MAX_CLOUD_COVER_TILE). If FILE_TILE is used, only the indicated tiles will be output.

Below are a few examples of commonly used projections. Fig. 13 depicts an Albers Equal Area projection, which is often used for the Continental Unites States (CONUS), due to the predominant East-West extent. Fig. 14 depicts a Lambert Azimuthal Equal Area projection, which is often used for areas with equal extent in all directions (e.g. Europe). Note that parallels are projected as circles, thus, a point that is very far away from the origin can have the same x-coordinate as a point in the middle of the coordinate system. This is amplified towards high latitude. Therefore, it is good practice to use a tile origin that is not too far away from the origin and preferably also well away from the poles (e.g. a tile origin at 90N/0E would result in non-intuitive behavior as it would be projected at 90N/10E). Fig. 15 depicts a UTM projection. Note that cylindrical projections are only valid for the area covered by

[2] Snyder, J.P. 1987. *Map projections--A working manual*: USGPO.

the cylinder, as such, any tile origin outside of the colored area would be invalid and would result in a transformation error. In addition, the precision quickly degrades in x-direction, although it is very favorable for areas with high North-South extent.

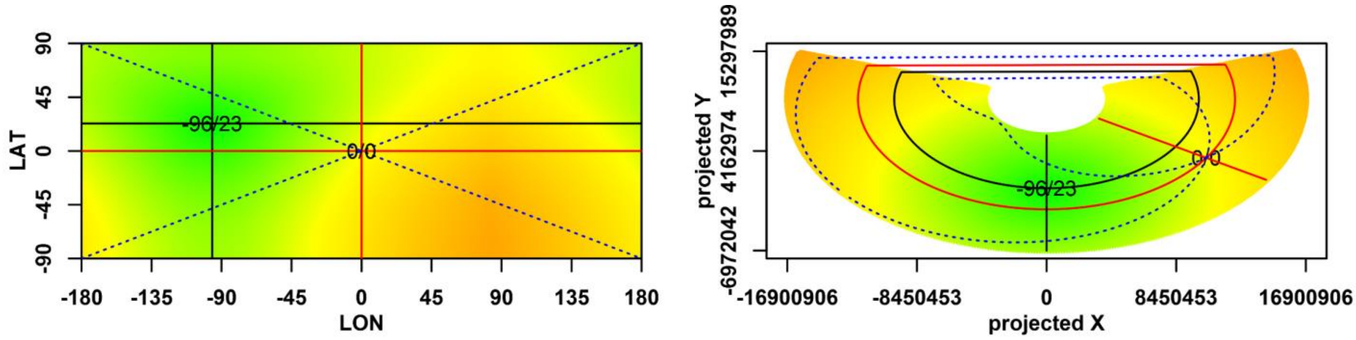

Fig. 13. Albers Equal Area projection example
center: -96W/23N, standard parallels: 29.5N/45.5N, datum: NAD83, ellipsoid: GRS80.

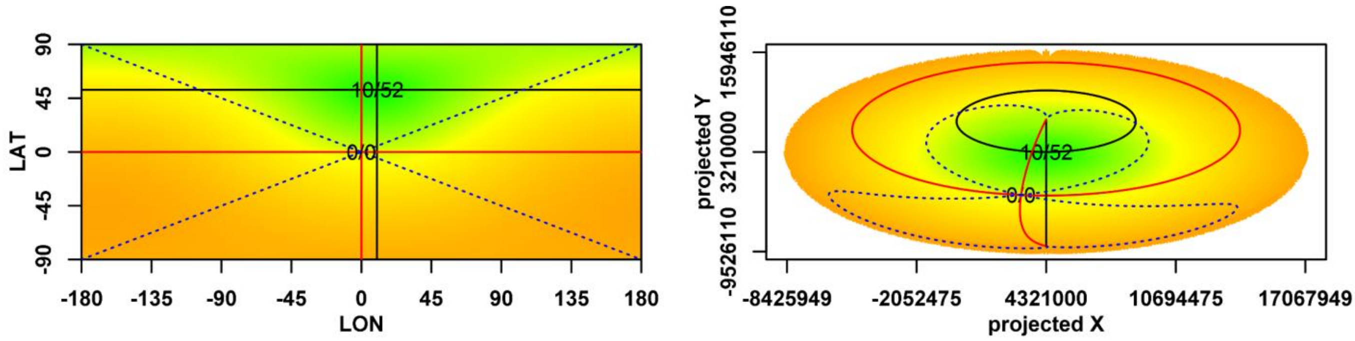

Fig. 14. Lambert Azimuthal Equal Area projection example
center: 10W/52N, false easting/northing: 4321000/3210000, datum: ETRS89, ellipsoid: GRS80.

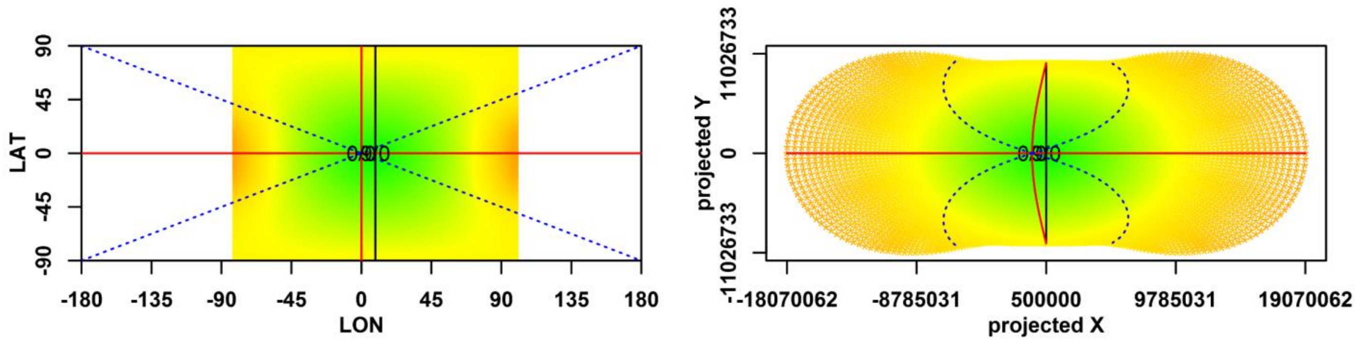

Fig. 15. Universal Transverse Mercator, zone 32 projection example
center: 9W/0N, false easting/northing: 500000/0, datum/ellipsoid: WGS84.

IX. Frequently Asked Questions

Something is wrong with the text files (including the parameter file), but I cannot see a mistake.

Make sure that the End-of-Line character is in Unix format (\n). Do not use the standard Windows text editors as they will automatically change EOL to Windows standard (\r\n).

One of the programs crashed.

Make sure to use full file names and avoid relative filenames containing characters like '.', '..', '~'; avoid special characters like spaces ' '. You may have found a bug…

The tile IDs of the processed Level 2 data have negative numbers, e.g. X-100_Y0100.

Make sure that the origin of the target grid (ORIGIN LAT / ORIGIN LON) is in the North-West of your study area.

Potentially, you have accidentally swapped latitude and longitude. Note that a geographic location in the North-West is not necessarily North-West in the output coordinate system, too (for an example see Fig. 14). Although not recommended, higher-level FORCE functions should be able to digest negative tile numbers (note that we did not test this exhaustively).

Is it possible to have a look at all the temporary layers that are created in the L2PS internals?

Theoretically yes, but this option should only be used by experts. You can re-compile the code in DEBUG mode, which features extensive output where images for most processing steps are saved. Note that these images are intended for software development and do not necessarily have intuitive file names; metadata or projections are also not appended. If DEBUG is activated, *force-level2* does not allow you to process multiple images or to use parallel processing (your system will be unresponsive because too much data is simultaneously written to the disc, and parallel calls to *force-l2ps* would overwrite the debugging images). For debugging, follow the steps summarized on page 23.

Transformation failed.Computing tile origin in dst_srs failed. Error in geometric module

Following error appears (L2PS) in the Level 2 logfile: Transformation failed. Computing tile origin in dst_srs failed...

This is most probably due to a bug in *force-parameter-level2* (see VI.B.1), which can be solved by adding following line before the projection definition:

PROJECTION =

Following error appears (L2PS): 'grep: QUEUED: No such file or directory. No images in …

Most probably, you have pasted the content of the file queue into the FILE_QUEUE parameter in the parameter file. FILE_QUEUE expects the file name (full path) of the file queue.

Following error appears (L2PS) in the Level 2 logfile: Unacceptable tier.

The acceptable tier level can be specified in the parameter file. Landsat Level-1T data (tier = 1) feature the best geometric correction and geolocation. We recommend to not use Level-1G or Level-1Gt images (tier = 2), unless you know what you are doing. See section VII.A.

Following error appears (L2PS) in the Level 2 logfile: DEM out of bounds

There is a problem with the DEM. The DEM must be provided in meters a.s.l., and nodata needs to be -32767. There is a security query, which ensures that the DEM is within -500 m and +9000 m.

Following error appears (L2PS) in the logfile: zero-filled temperature.

Our Landsat cloud detection is in need of temperature data. The Landsat-8 TIRS was reconfigured due to anomalous current levels, and for a certain period of time, data were distributed by USGS with zero-filled temperature data. The USGS has fixed the problem and the data are/were reprocessed in a phased processing strategy. You should regularly check on updates from USGS, re-download the failed images and process them again. This might happen again, though.

Following error appears (L2PS) in the logfile: Unable to open MTL file. Parsing metadata failed.

FORCE can only handle Landsat data processed by the Level 1 Product Generation System (LPGS) of USGS, which comes with an MTL text file (contains the metadata). Data processed with the outdated National Land Archive Production System (NLAPS) are not supported. Sorry.

Following error appears (L2PS) in the logfile: tar.gz container is corrupt.

There are two possible reasons: 1) the file downloaded from USGS is corrupt, incomplete, etc. In this case, delete the image, remove it from the file queue and download/process again. 2) force-level2 checks for non-zero exit code when extracting the images. On some systems, the tar/gzip programs throws a warning each time it extracts an image; this is probably related to some write permissions or mount settings. There is not much to do about this from our side. You need to fix your settings, speak with your admin. Alternatively, you can disable the exit-code check by changing, removing or commenting following lines in *bash/force-level2.sh*. If doing this, follow-up errors will occur if there really was a problem with the file.

```
if [ ! $? -eq 0 ]; then
  echo "$BASE: tar.gz container is corrupt."
  FAIL=1
Fi
```

An error like this appears (L2PS) in the logfile: L1C_T21MXM_A007643_null: unknown Satellite Mission. Parsing metadata failed.

This can happen when Sentinel-2 data downloads are incomplete. Delete the image, remove it from the file queue and download/process again.

Following error appears (L2PS) in the logfile: L2PS is already running. Exit.

FORCE L2PS has a built-in safeguard, which was implemented to allow safe operational and scheduled processing. FORCE L1AS and FORCE L2PS can be used for NRT processing, i.e. data can be downloaded and processed with n CPUs at given intervals. As the processing can take longer than these intervals, the safeguard protects your system from launching another n processing jobs, which may exceed the N CPUs available on your machine. You can disable the safeguard by changing, removing or commenting following lines in *bash/force-level2.sh*:

```
# protect against multiple calls
if [ $(ps aux | grep 'L2PS' | wc -l) -gt 1 ]; then
  echo "L2PS is already running. Exit." > $OD/FORCE-L2PS_$TIME.log
  exit
fi
```

Following error appears (L2PS) in the logfile: Unable to lock file. Error in writing products! Tiling images failed! Error in geometric module.

There is a write problem. 1) If L2PS was aborted in a previous run, some left-over lockfiles might exist (*.lock). In this case, FORCE cannot lock the file as it is already ‘locked’. Temporary locking the files is important as we’ll have write conflicts from parallel calls if not doing this. You need to remove the lock files. 2) The lockfile generation timed out. This may happen if there is too much I/O activity on your system, such that FORCE is not allowed to write data for quite some time. Reduce I/O from other processes/users. Try to use fewer parallel processes. Try to increase the delay. Try writing to a disc that can handle the I/O, preferably directly attached to the server.

Following warning appears on the screen: ‘lockfile creation failed: exceeded maximum number of lock attempts’

There is a known problem with CIFS mounted network drives. You can ignore these warnings; they are no fatal errors. But you might want to inspect the file queue after Level 2 processing, as there is a minor possibility that there were some conflicts due to parallel write attempts: a few images might not have been switched from QUEUED to DONE status. This does not imply that the image was not processed (check the logfile as well).

There are holes in my processed Level 2 images. Why?

Nodata values in the DEM are masked. Impulse Noise is attempted to be detected and is masked out. The image border (including SLC-off stripes) is buffered by one pixel as these pixels are often erroneous. The masks are applied all output products.

The programs don’t run and there are strange symbols on the screen.

You have probably copied text from this document to your shell. This might be an encoding issue. Try to manually type the commands.

X. Known Issues

There are a couple of known (and surely unknown) issues in FORCE, which can affect image quality. Please note that FORCE was developed for scientific purposes to fulfill the research needs of the author team. FORCE started as a project to correct Landsat imagery in Southern Africa for land cover change research. In the meantime, it was further developed, new modules were added, support for Sentinel-2 was added and functionality for other environments were incorporated (e.g. AOD estimation over dense dark vegetation or improved topographic correction). FORCE was tested and validated in a variety of environments and settings, and was distributed in the hope that it will be helpful to fulfill your research needs, too. However, there are still a number of issues, which we didn't have time to solve yet. Note that some of the 'problems' mentioned below are just messages of caution; in any case, we advise you to make use of the quality flags. You are invited to develop solutions for the problems mentioned below, and to report other unknown issues, mistakes and bugs. I would welcome if you would pass back improvements, in which case I will try my best to find time to review and incorporate these changes into the main build.

David Frantz: david.frantz@geo.hu-berlin.de

1) AOD estimation in bright landscapes

FORCE L2PS estimates AOD over dark targets. Thus, if there are none or few, AOD estimation becomes less reliable or might fail completely. If there is not a single valid dark target, a global fallback value is used, which might not represent the actual conditions very well. Use these scenes with caution, and take a look at the AOD quality flags.

You can use externally provided fallback values to counter this.

Note that water bodies might not be suitable per se; if they are very turbid, they can even be brighter than the land surface in the visible bands; AOD cannot be estimated from these targets.

Note that the presence of vegetation is also not sufficient per se; the vegetation needs to be dark and dense.

2) AOD estimation over the ocean

AOD estimation over the ocean works well in general. However, white-caps are problematic as they increase water reflectance, which yields higher AOD values. A test is made if the water body is unrealistically bright, in these cases the water body is rejected; this may happen over large lakes or ocean.

3) AOD estimation in mountainous areas

AOD estimation in mountainous areas is still problematic. AOD is estimated for each suitable dark target, then scaled to a reference elevation using an exponential model, then averaged for coarse resolution grid cells of 5 km, and then interpolated to generate an AOD map. Finally, AOD is scaled back to the individual pixel's elevation. This works fine if the elevation range is moderate and if the assumed model exponential matches the actual conditions. On the contrary, it may give unrealistic AOD in high mountains (if the exponential model doesn't represent the actual conditions). This is further aggravated by the fact that mountains are often rock formations, are covered in snow or are clouded, which reduces the number of potential targets for AOD estimation (thus more interpolation is needed). If this happens, estimated AOD will be too high, and the spectra look 'skewed' and are often negative in the shortest wavelengths. Use these scenes with caution, and take a look at the AOD quality flags, and the sub-zero flag.

4) Environment correction

The environment corrects for adjacency effects, i.e. it removes the part of the radiation that comes from neighboring surface elements. In some cases, the correction may be too strong. The result can be sub-zero pixels, e.g. in the visible wavelengths. It was mainly observed over darker-than-usual pixels, e.g. topographic cast shadows, cloud shadows, or forest shadows (because the environment reflectance is estimated over un-shaded pixels). Because of spatial resolution, it is more pronounced in Sentinel-2 imagery. Environment correction will also aggravate problems caused by overestimated AOD (e.g. issue # 3). Be sure to check the sub-zero flag when using this option.

5) Topographic correction of poorly illuminated areas

The implemented topographic correction is not intended to correct hard shadows. Areas with illumination angle > 80° are not expected to be corrected reliably. Areas with illumination angle > 90° (self-shadow) are not corrected. Make sure to look at the illumination flags.

6) 'Flat' vegetation spectra

Flat vegetation spectra (in the visible bands) were observed under low sun elevations. This may be related to radiative transfer (is based on 5S), AOD overestimation or to some bug. We could not yet confirm the reason for this.