Using JASMIN/CEMS for the Generation of ECV Data Records in ESA CCI and EC Projects

O. Danne\(^1\), T. Block\(^1\), T. Storm\(^1\), C. Brockmann\(^1\), S. Kharbouche\(^2\), J.-P. Muller\(^2\), S. Blessing\(^3\), R. Giering\(^3\)

\(^1\)Brockmann Consult, \(^2\)UCL/MSSL, \(^3\)FastOpt GmbH

JASMIN Conference, RAL, Harwell/UK, 27-28 June 2017
Presentation Overview

• ECV overview

• The Projects:
  • QA4ECV (EU)
  • SST-CCI (ESA)
  • FIDUCEO (EU)
  • FIRE-CCI (ESA)

and how we use JASMIN/CEMS for them

• QA4ECV: Technical aspects and challenges
ECV Overview

• GCOS (Global Climate Observing System) list of 2010: 50 Essential Climate Variables (ECVs)

• All ECVs are technically and economically feasible for systematic observation

• International exchange is required for both current and historical observations
## ECV Overview

<table>
<thead>
<tr>
<th>Domain</th>
<th>GCOS Essential Climate Variables</th>
</tr>
</thead>
</table>
| **Atmospheric** (over land, sea and ice) | **Surface:**[1] Air temperature, Wind speed and direction, Water vapour, Pressure, Precipitation, Surface radiation budget.  
    **Composition:** Carbon dioxide, Methane, and other long-lived greenhouse gases[3], Ozone and Aerosol, supported by their precursors[4]. |
| **Oceanic**                 | **Surface:**[5] Sea-surface temperature, Sea-surface salinity, Sea level, Sea state, Sea ice, Surface current, Ocean colour, Carbon dioxide partial pressure, Ocean acidity, Phytoplankton.  
    **Sub-surface:** Temperature, Salinity, Current, Nutrients, Carbon dioxide partial pressure, Ocean acidity, Oxygen, Tracers. |
| **Terrestrial**             | River discharge, Water use, Groundwater, Lakes, Snow cover, Glaciers and ice caps, Ice sheets, Permafrost, Albedo, Land cover (including vegetation type), Fraction of absorbed photosynthetically active radiation (FAPAR), Leaf area index (LAI), Above-ground biomass, Soil carbon, Fire disturbance, Soil moisture. |
QA4ECV (Quality Assurance for Essential Climate Variables):

• A project funded under theme 9 (Space) of the Framework Program 7 of the European Union

• The project will
  • develop a Quality Assurance (QA) system for observational data products
  • perform a detailed practical demonstration of the QA concept concentrating on six Essential Climate Variables (terrestrial and atmosphere)

• ECVs:
  • Atmosphere: nitrogen dioxide (NO₂), formaldehyde (HCHO), carbon monoxide (CO)
  • Terrestrial: albedo, fAPAR, LAI. All including per-pixel uncertainties.
Surface albedo algorithm overview

- Spectral and broadband albedo retrievals, daily, global at 0.5° and 0.05°:
  - MODIS wavelengths
  - HCHO, NO2 and CO wavelengths
  - Broadband (0.4-0.7μm, 0.7-3.0 μm, 0.4-3.0μm)

- Combination of sensors onboard GEO and LEO satellites, 1981-2016

- Requires spectral mapping to target wavelengths
  (→ S. Kharbouche, UCL/MSSL, JASMIN conference 2016)

- Use optimal estimation algorithm for per-pixel retrieval including uncertainties

- Uses MODIS collection 6 priors derived from 16 years of MCD43A1,2 (500m)
Scientific processing flow for surface albedo

1. **VGT Level-1p**
   - TOA Reflectance
   - Pixel Identification
   - Atmospheric Correction
   - BRF
   - Wavelength Transformations + Broadband Integration

2. **MERIS Level-1b**
   - MODIS Prior
   - BBDRs
   - BEAM Binning
   - BRDF Kernel Computation
   - Upscaling using Energy Conservation
   - Albedo Integration
   - GlobAlbedo Product

**Key**
- Main Process
- Data/Product
- Click to see process
- Click to see more details
- Click to return to main chain
Example result: BHR broadband albedo, Jan 1\textsuperscript{st}, 2008, 0.05\degree
Processing of surface albedo: System setup at CEMS

- BEAM (SNAP) Toolbox core software (Java)
- BEAM (SNAP) reader and writer plug-ins (Java)
- BEAM (SNAP) Graph Processing Framework for scientific processors
- QA4ECV Albedo plug-ins (Java)
- PMonitor (Python) for supervision of LSF job submission, logging, reporting (BC in-house development)
- ‘Staging’ tools for mosaicking, quicklook and movie generation (developed by S. Kharbouche, UCL/MSSL, using Java/Python/ffmpeg)
Process for surface albedo

1. Job Monitoring
   - PMonitor
   - Python Script

2. Job(s) Definition
   - Bash Shell Script

3. Job(s) Registration
   - LSF Job Submission

4. BEAM Process Definition
   - Bash Shell Script

5. BEAM GPF call
   - Running Java processor

6. BEAM Target Product

PMonitor basics:

• PMonitor is a Python-based workflow engine for scripting solutions. It is used as "client software" to control bulk production.

• The main PMonitor Python module handles tasks and their dependencies (as formal inputs and outputs) and executes them on a thread pool. Most importantly, it provides
  • a thread pool with a task queue of mature tasks with inputs available
  • a backlog of tasks with inputs not yet available
  • a report file that records all completed calls and the paths of output product (set) names
QA4ECV - Albedo

Albedo processing challenges:

• Most processing steps are straightforward, as they have completely independent pairs of input/output. E.g., daily processing over 1 year can be fully parallelized into 365 parallel LSF jobs

• Usually, the LSF default job time and memory settings are sufficient

• However, due to the underlying algorithm, the ‘optimal estimation’ has more complex requirements:
  • Overall ~4 million single LSF jobs for ~33 years of input data
  • To generate an albedo product for a given day, a very large number of BBDR input products from an 18-month time window is used → heavy data I/O
  • Difficult to find a balance between
    • holding data in memory, or
    • write intermediate data products, or
    • discard intermediate data and repeat computations
Albedo processing challenges (2):

- Multi-step retrieval → the sequence of jobs being processed is not arbitrary
- PMonitor: frequent supervision if required intermediate products are available

- Specific jobs require more memory/time than defaults (i.e. generation of high-resolution global albedo mosaics)

- Thanks to CEDA, we could benefit from a couple of high-priority processing phases with a certain number of cores dedicated to the QA4ECV project

- The processing requirements for the project can be met with the current setup.

- We appreciate further hints for improvements, but would also like to thank CEDA helpdesk (Fatima and colleagues) for all help we got so far!
The Two-Stream Inversion Package (TIP) - an ECV algorithm for fAPAR & LAI derived from albedo

1) Forward model
2) Inversion (TIP)
3) Processing chain
4) Performance on CEMS
5) Results

S. Blessing,
R. Giering
(FastOpt)

Forward model (Two-Stream):
inputs:
- effective Leaf Area Index (LAI),
- effective leaf reflectance,
- effective leaf transmittance
- soil or background albedo

outputs:
radiant fluxes in the canopy:
- reflected,
- transmitted,
- and absorbed (→ fAPAR).

S. Blessing, R. Giering (FastOpt)
The Two-Stream Inversion Package

**Optimisation:**
- find 2-stream parameters matching BHR-VIS and BHR-NIR

**At Optimum:**
- optimal LAI
- simulated VIS-absorption in canopy is fAPAR

**Error Propagation:**
- use 2nd derivative at optimum to determine posterior covariances

**From Covariances:**
- uncertainties on LAI and fAPAR

QA4ECV - fAPAR & LAI
QA4ECV - fAPAR & LAI

Processing flow for TIP FAPAR & LAI
Results for TIP LAI/fAPAR

- Recent QA4ECV processing of daily 5km resolution BHR albedos took ~10 days on CEMS for whole period of ~33 years

- per pixel uncertainty is provided from propagation of full BHR covariance information

- output is NetCDF-4 classic with internal compression, conforming to CF conventions and CCI requirements
**ESA SST-CCI:**

- A project funded by the European Space Agency (ESA) as part of the Agency’s Climate Change Initiative (CCI) Programme
- Focus on various issues related to Sea Surface Temperature (SST), e.g.:
  - analysis of scientific requirements relating to climate
  - algorithm selection and development
  - production and validation of SST datasets
- ECVs: Sea Surface Temperature (SST)

**FIDUCEO (FiDelity and Uncertainty in Climate data records from Earth Observations):**

- A project funded by the Horizon 2020 Framework Programme of the European Union
- The project will produce new climate data records with complete and traceable estimates of stability and uncertainty
- ECVs: SST, surface albedo, atmospheric water vapour, aerosol optical depth
  - Both projects require the generation of *satellite-to-satellite data matchups* for various purposes.
  - Brockmann Consult responsible: T. Block

JASMIN Conference, RAL, Harwell/UK, 27-28 June 2017
Satellite-to-satellite data matchups

• Purpose:
  • Sensor cross-calibration
  • Harmonisation of multi-decadal measurement time-series
  • Validation of retrieval algorithms

• Definition „matchup“:
  • Two sensors view the same environmental scene simultaneously
  • View: search for pairs of sensor pixels (whole radiance spectrum)
  • Simultaneously: pixels are close in a geodesic sense (below xx km distance)
  • Simultaneously: pixels are acquired close in a time sense (below xx secs distance)

• Additional conditions:
  • Only land/water pixel
  • Cloud coverage
  • Viewing angles, etc ...
• Numerically intensive process
  – Search in large satellite datasets (> 150 TB)
  – Parallel approach

• Three stage process
  – Extract satellite metadata (location, time, ...) and store to database
  – Coarse search (detect geometric/time intersections)
  – Fine search (identify matching pixels within intersecting geometries)
• **Extract Metadata:**
  
  – Bounding polygon of product from geolocation (grey)
  
  – Time axis (green). Allows estimation of acquisition time in lon/lat space
  
  – Other metadata:
    
    • Sensing time
    • Orbit node
    • Processing version
    • Etc ...
• Coarse search:
  – Intersect geometries
  – Use time axis to calculate overflight times
  – Decide if geometry contains possible matchups
• Fine search:
  – For each geometric intersection:
    – Find closest matching pixels
    – Run further conditions on set of matchups
• Result data
  – stored in NetCDF file (MMD – multisensor matchup dataset)
  – Self contained format, all input sensor variable data in mxn window around matchup

• Processing on CEMS
  – 1.2 Million metadata records (referencing ~120 TB), 7 sensors, 23 platforms
  – Performing e.g. AVHRR N17/N18 in 2:30 hrs using 72 nodes (5.5 years time series)
ESA Fire-CCI:

- A project funded by the European Space Agency (ESA) as part of the Agency’s Climate Change Initiative (CCI) Programme
- This project
  - aims to improve consistency of burned area products
  - shall improve algorithms for both pre-processing and burned area detection
  - incorporates error characterisation
- The goal is to produce multiple global long-term time series of Burned Area maps
- Two burned area products will be delivered:
  - Pixel product (monthly, ~300m)
  - Grid product (0.5°)
- Brockmann Consult responsible: T. Storm
Project requirements and challenges:

- process a very large dataset:
  - MODIS TERRA L2 archive (MOD09GA + MOD09GQ)
  - global land data, 2000-2016
  - 450 TB

- generally: high download and processing performance needed
- simple usability and possibility to employ generic processors
- sufficient storage space
  - not needed to store all input data at once
  - but considerable fraction needed for processing

→ JASMIN/CEMS facilities have been identified as infrastructure capable of fulfilling these project requirements
ESA Fire CCI

Project status:

• downloading has started in January

• processing has started in March
  • processor: burned area retrieval, written in Python by University of Alcala, Spain
  • monitoring and scheduling done using PMonitor software, developed by Brockmann Consult
  • had two months of integration and evaluation

• started to process actual dataset at end of May

• full time series was completed on June 9th for African land mass

• currently: processing Asian land mass
Burned area in central Africa, 1st half of Jan 2010
Red = more burned area

Burned area in central Africa, 2nd half of Jan 2010
Red = more burned area
Thank you!