Delivering MACC data
What have we learned?

Miha Razinger

Cihan Sahin, Richard Engelen and Xiaobo Yang

Copernicus Climate Data Store workshop
3-6 March 2015
MACC-III is the precursor of the Copernicus Atmosphere Monitoring Service and it’s the fourth in a series of FP6, FP7 and Horizon 2020 EU R&D projects (since 2005).

The main aim of the series of projects was to develop pre-operational services in the wider field of atmospheric composition, which meet the needs of users.

It is coordinated by ECMWF and the consortium comprises 36 partners from 13 countries.
From EO to policy-relevant products

Over 60 EO instruments are assimilated in the global system.

Boundary conditions feed an ensemble of high-resolution European AQ systems (in order to assess uncertainties).

More data are assimilated (in particular hourly surface AQ concentrated by EEA/EIONET).

Policy-relevant (here health indicator for ozone) products are delivered. They are “maps with no gaps”, which observations alone don’t provide and are essential to assess impacts.
European Centre for Medium-Range Weather Forecasts

Radiation and ozone layer

Surface fluxes: greenhouse gases, fires, emissions (GFAS, MACCity, MACC/TNO)

Global atmospheric composition

European Air Quality

http://atmosphere.copernicus.eu
Power users - routine use of MACC-II products as part of time-critical activities (limited area AQ modellers, satellite agencies, companies producing public AQ and UVI forecasts, air quality apps developers, solar energy sector ...).

<table>
<thead>
<tr>
<th>Main products categories</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global NRT Analyses &amp; Forecasts</td>
<td>150 daily users</td>
</tr>
<tr>
<td>Regional NRT Analyses &amp; Forecasts</td>
<td>125 daily users</td>
</tr>
<tr>
<td>MACC Global Reanalysis 2003-2012</td>
<td>1600 registered users</td>
</tr>
<tr>
<td>Solar Radiation</td>
<td>110 users, ~20000 requests/yr</td>
</tr>
</tbody>
</table>
Copernicus Atmosphere Monitoring Service

- Operational delivery of atmospheric composition services
- Global and European regional scale
- Initial period from 2015 – 2020
- ECMWF is in charge of implementation

in total **68** products and services
Anticipating questions from potential users ...
What products and services can I get from MACC?  

**single entry point**

How can I access the data? Where can I find a service?  

**it depends**

How can I read and interpret the data? How can I use a service?  

**it depends**

Where to get information about the quality of a product?  

**it depends**

Is a product suitable for operational usage?  

**it depends**

What am I allowed to do with the data?  

**single data licence**
Diversity of data producers, number of data formats, large variability of dataset volumes, existing data serving methods
MACC product catalogue

• Start small: comprehensive and up-to-date inventory
• **150** individual products and services accompanied with description, geographical and temporal coverage and links to access the data, to data browsing page, documentation, validation reports, contact points ...
• Able to export INSPIRE and WMO Core compliant meta data
• Will be used as a source for Catalogue Service for the Web (CSW)
<table>
<thead>
<tr>
<th>Product</th>
<th>Name</th>
<th>Service Type</th>
<th>Product Family</th>
<th>Parameter</th>
<th>Service Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>European air quality assessment report 2010</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>NO2, Birch pollen, PM2.5, O3, PM10</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>European-scale AQ ozone forecast by SILAM</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>O3</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>Green scenarios</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>NO2, PM2.5, O3, PM10</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>MACC-IFS-MOZ reanalysis of global carbon monoxide</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>CO</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>MACC-IFS-MOZ reanalysis of global formaldehyde</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>HCHO</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>MACC-IFS-MOZ reanalysis of global ozone</td>
<td>Ozone and Ultraviolet radiation, Air quality and atmospheric composition</td>
<td>Reactive gas, Greenhouse gas</td>
<td>O3</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>MACC-IFS-MOZ reanalysis of global reactive nitrogen oxides</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>NOx</td>
<td>PRE-OP</td>
</tr>
<tr>
<td></td>
<td>MACC-IFS-MOZ reanalysis of global sulphur dioxide</td>
<td>Air quality and atmospheric composition</td>
<td>Reactive gas</td>
<td>SO2</td>
<td>PRE-OP</td>
</tr>
</tbody>
</table>

Please use the search criteria on the left to filter products. Once you get a list of products, you may check a product's details by clicking on it. You can also type one or more keywords to search the catalogue.

Explanation of Service Status levels.

If you have any comments, please feel free to contact us using this form.

http://atmosphere.copernicus.eu/catalogue
MACC-IYS-MOZ reanalysis of global carbon monoxide

Description:
This service provides global reanalysis of carbon monoxide using the chemistry of the MOZART model.

Service type: Air quality and atmospheric composition
Product family: Reactive gas
Parameter: CO
Geographical area: Global (-180, 180, -90, 90)

Potential temperature; Potential vorticity; Model; Surface; Pressure
Time resolution: Monthly; 3-hourly
Datatype: Reanalysis
Service status: Pre-operational

Links: Plots, Data access, Verification results, Validation reports, Contact us, XML
Validation activities

Verification of Regional Services

Validation Reports

The regional modelling systems are being validated on a 3-monthly basis. The validation reports are available here:

- Validation reports

Quick-look verification pages

The verification of the MACC-II Regional Air Quality services over Europe is based on comparisons with in-situ surface observations of the following air pollutants: ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide and PM10 aerosols. These hourly observational data are delivered to the MACC-II project close to real-time. They are preliminary and not validated and can therefore not be used for checking compliance with air quality regulations or for any purpose other than the evaluation of MACC-II Regional Air Quality products. Detailed information can be obtained from the data owners.

Forecast verification against in-situ observations

Maps of the forecasts of surface pollutants (ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide and PM10 aerosols) produced by each of the 7 individual models or from the ensemble on a 3-hourly basis, overlaid with in-situ observations as coloured dots. Available for the past 30 days.

Analysis verification against in-situ observations

Maps of the ozone analysis at the surface produced by each of the 7 individual models or from the ensemble on a 3-hourly basis, overlaid with in-situ observations as coloured dots. Available for the past 30 days.
Distributed data access

- FTP servers
- WCS servers
- ECPDS service
- ECMWF data servers
- WMS servers
- API services
- Other

Product catalogue
Two examples of delivery systems for the MACC global datasets:

1. ECMWF data servers
2. ECPDS data delivery
Delivering global model results

- Hosting the global near-real-time and the reanalysis dataset
- Very capable of serving big volumes of data
- Services build on top of existing and upcoming ECMWF systems
- Most of the data is archived in MARS
- Data volumes: reanalysis 50 GB/day, 22 TB in total, NRT dataset 200 GB/day, 2.5 years, 24 TB
- Not best suitable for data browsing and occasional / light-weight data usage
### MACC Reanalysis

**Select date**

- Select a date in the interval 2003-01-01 to 2012-12-31
- Start date: 2003-01-01  End date: 2012-12-31
- Reset

**Select a list of months**

<table>
<thead>
<tr>
<th>Year</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Select time**

- 00:00:00
- 06:00:00
- 12:00:00
- 18:00:00
- Select All or Clear

**Select step**

- 0
- 3
- 6
- 9
- 12
- 15
- 18
- 21
- 24
- Select All or Clear

**Select level and parameter**

| Parameter          | 1000 | 925 | 850 | 775 | 700 | 600 | 500 | 400 | 300 | 250 | 200 | 150 | 100 | 70  | 50  | 30  | 20  | 10  | 7   | 5   | 3   | 2   | 1   |
|--------------------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Carbon monoxide    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Divergence         |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Formaldehyde       |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| GEMS Ozone         |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Geopotential       |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
ecCharts components

Highly available end-to-end service is achieved by:
- Load balancing
- Distributed architecture
- Virtualization
- Service Oriented Architecture (SOA)
ECMWF data servers

- Frontend: JQuery & AJAX based web application
- Backend: Django web framework
- Catalogue of products & User preferences: MongoDB collections of JSON documents
- Distributed Object Caching: Memcached
- SOA components: Bespoke python based framework (with twisted)
- SOA services: mostly in Python and C.
- Data: From MARS (Retrieve to ingest on ecCharts data clusters)
- Data cluster: Data replicated on standard Unix file system.
- Data access: Based on MARS language
- Data related operations: GRIB API
ECPDS data service

- Serves NRT global, global BC and NRT GFAS data streams
- FTP-push and –pull service
- suitable for large data transfers
- 24/7 monitored and supported
- High availability
- Load balance
- Per-dataset, per-user quotas
- Scalable
- User Account Management
- Monitoring Facilities
Successes

- Large and growing user base, in general high level of user satisfaction
- Multi-purpose catalogue: data discovery, MD export, portfolio documents, CSW
- Manage to remain flexible when dealing with a diverse set of datasets and user requirements
- Use of JIRA issue tracking system
Some lessons

• Prescribe (and police) format specification and provide data format validators. Distribute sample code if you want to be certain

• Provide better subsetting services to reduce volume of data that needs to be transferred (subareas, timeseries, vertical profiles ...)

• Value (and headache) of experimental datasets

• Exposing catalogue to web crawlers
Website: http://atmosphere.copernicus.eu
Contact: info@copernicus-atmosphere.eu