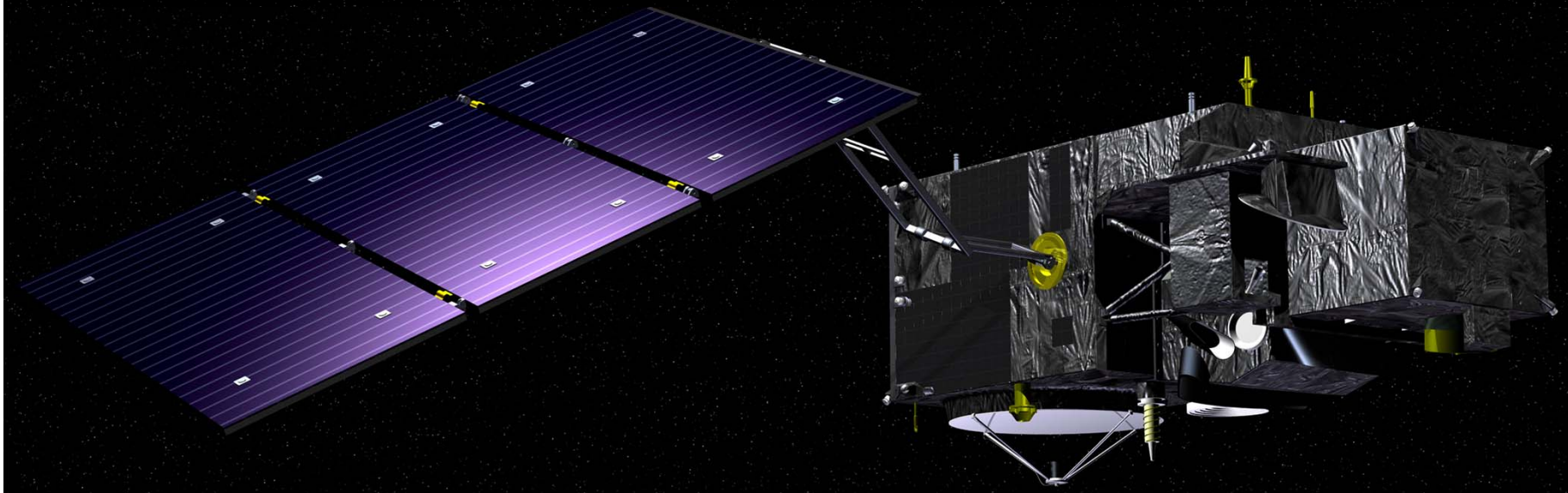
On data quality in the generation of products, tools, and services



Bojan R. Bojkov

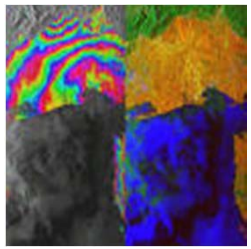
Head, Sensor Performance, Products and Algorithms

Directorate of Earth Observation Programmes

European Space Agency (ESA/ESRIN)

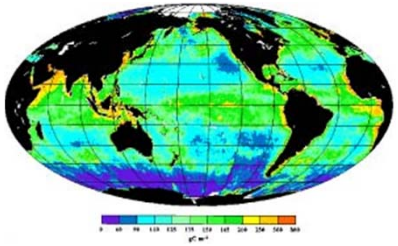
Sentinel-3

Copernicus Sentinel missions



Sentinel 1 (2013) – SAR imaging

- *All weather, day/night applications, interferometry*



Sentinel 2 (2014) – Superspectral imaging

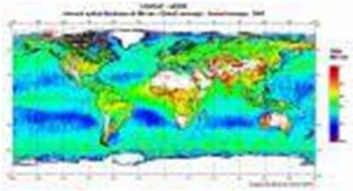
- *vegetation, forestry, security*

Sentinel 3 (2014) – Ocean/land monitoring

- *Wide-swath ocean color and surface temperature sensors, altimeter*

Sentinel 4 (2019) – Geostationary atmospheric

- *Atmospheric composition monitoring, trans-boundary pollution*



Sentinel 5p/5 (2015/2020) – Low-orbit atmospheric

- *Atmospheric composition monitoring*

Copernicus Sentinel missions (ii)

Fact: in the Copernicus era numerous products/tools/services (in addition to the core services) will be developed – some by combining very different datasets and others far-removed from the EO-domain and the EO data originators

but how do we ensure that the datasets, EO and non-EO, used in the generation of a product/tool/service are actually suitable for use?

Example: an urban air quality health risk assessment tool

The City of Ottawa estimates ~500 premature deaths/year attributed to air quality (capital area: ~1.5M inhabitants)

The ESA Data User Element supported the "*Decision Analysis Service for Urban Air Pollution Health Risk Assessments*" project (http://due.esrin.esa.int/prjs/prjs106.php) in 2009-2010

The project combined satellite air quality data, in-situ measurements, local model for air quality, transportation and health risk valuation to provide a tool to analyse the impact of air quality on health for the City of Ottawa

A-MAPS Environmental

Risk Sciences INTERNATIONAL

Ottawa

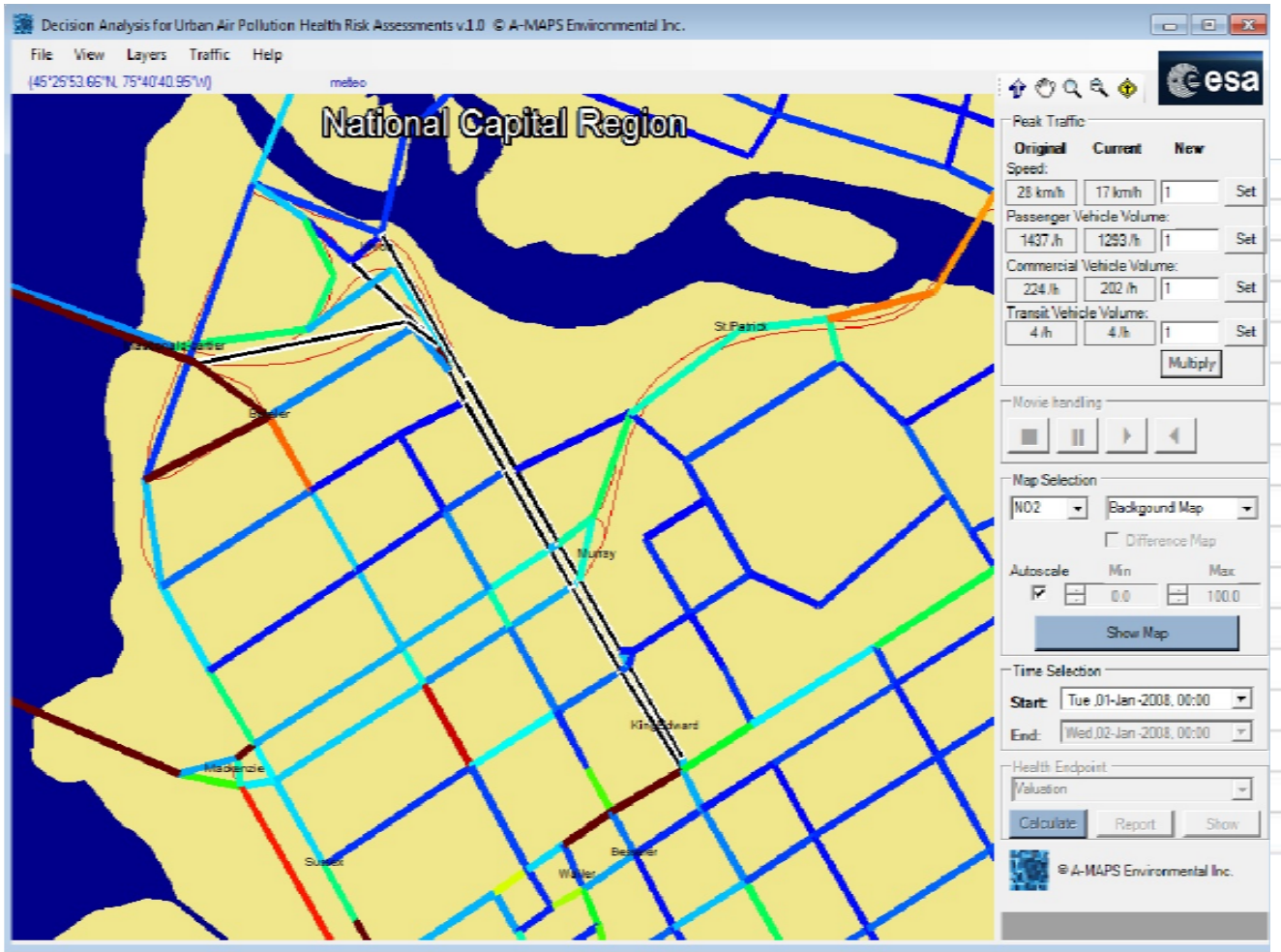
Air Quality impact in downtown Ottawa

20% traffic reduction on a main artery

Traffic density

PM 2.5 impact

Health cost savings



But how do we know this is correct?

- The premise of the **QA4EO** framework is to facilitate interoperability and harmonisation across all Social Benefit Areas, **in terms of data**. QA4EO is dependent on the implementation of the following principles:
    - *Accessibility / Availability*
    - *Suitability / Reliability*

  and both requiring their effective communication to all stakeholders

- **QA4EO is a common sense approach** of documenting the generation, validation, limitations/applicability, etc. of an EO dataset
  ➔ *used since the first EO mission in 1960!*

- QA4EO documentation is available at http://www.qa4eo.org/

So in the Ottawa project:

- **Air Quality Health Index (AQHI) and Air Quality Benefits Assessment Tool (AQBAT)** - Judek S., Stieb D., Jovic B. 2006. Air Quality Benefits Assessment Tool (AQBAT) release 1.0. Ottawa: Health Canada

- **Satellite data** - product specifications and limitations reports available from the space agencies, for example for the aerosol and NO_2 data used in the project (http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/documents/v003)

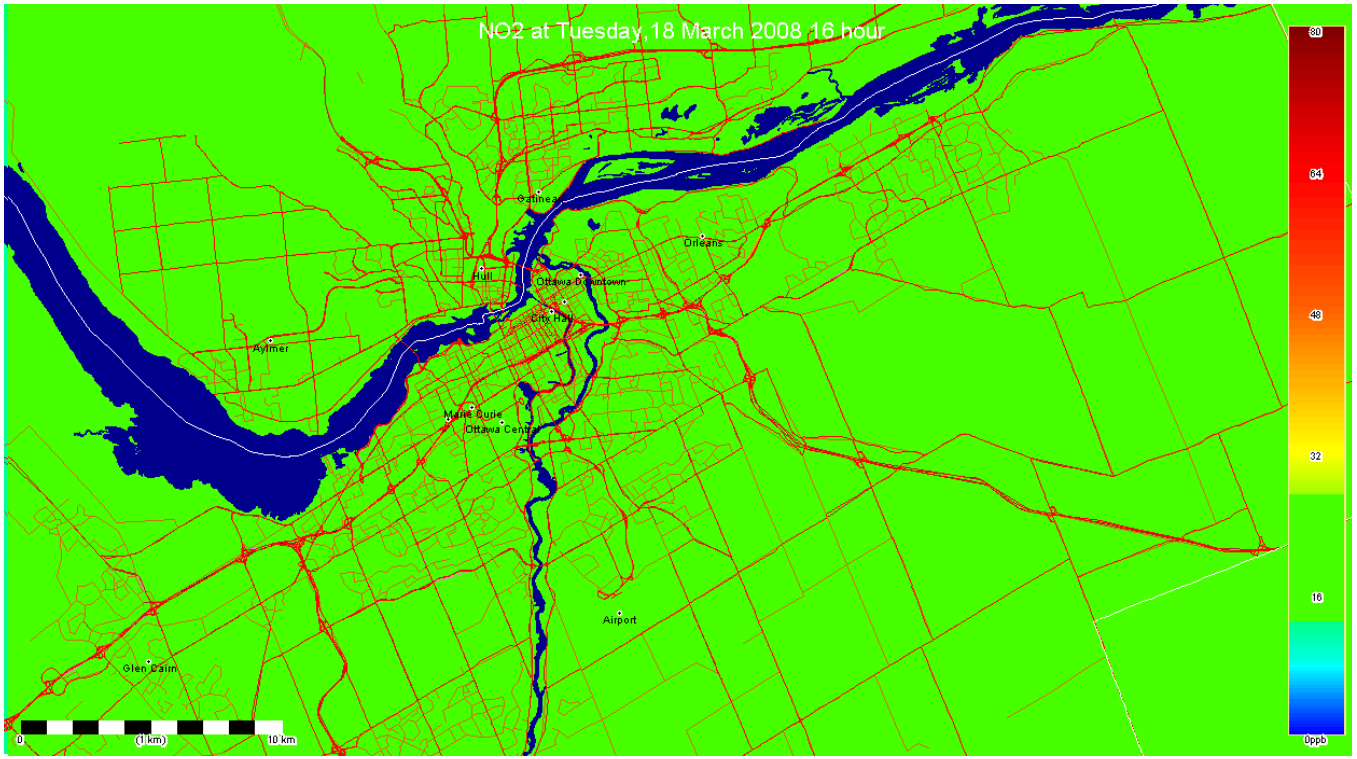
- Similar source information (and documentation) available for the **Ottawa traffic model**

But what about the suitability of the air quality model (with sat.) prior to input into the AQHI and the AQBAT?

1 - Did satellite data help the model?

The satellite data, of known quality, was used as boundary conditions, improving the representativeness urban air quality model

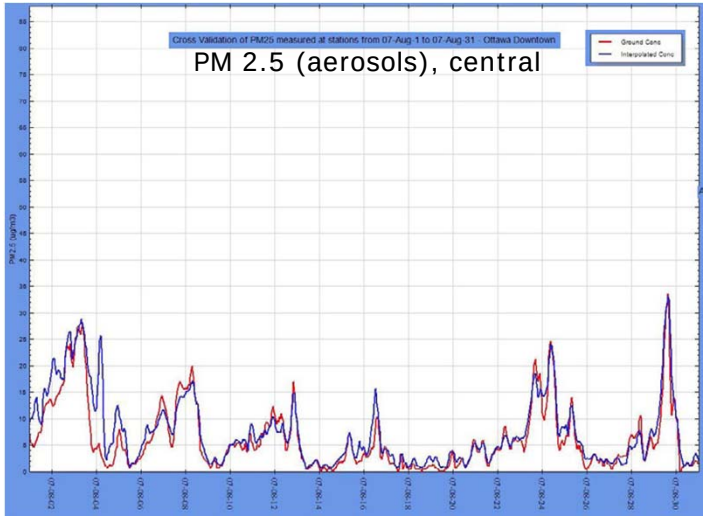
No satellite data

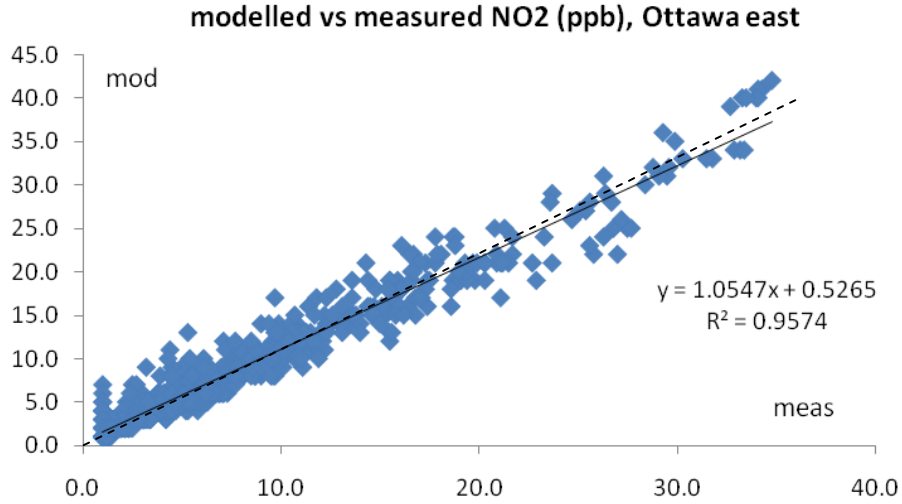


NO_2 model map generated using 8 ground-stations

2 - The AQ model required validation…

Extensive 2-year comparisons of the urban AQ model (initialised with satellite data) was performed with measurements from 8 fixed and a mobile air quality measurement unit, of documented quality, from the City of Ottawa **using best practise measurement and comparison protocols**





The inter-comparison analysis thoroughly reviewed for temporal and seasonal representativeness and the results were documented (see project validation documentation at http://due.esrin.esa.int/prjs/prjs106.php)

In conclusion

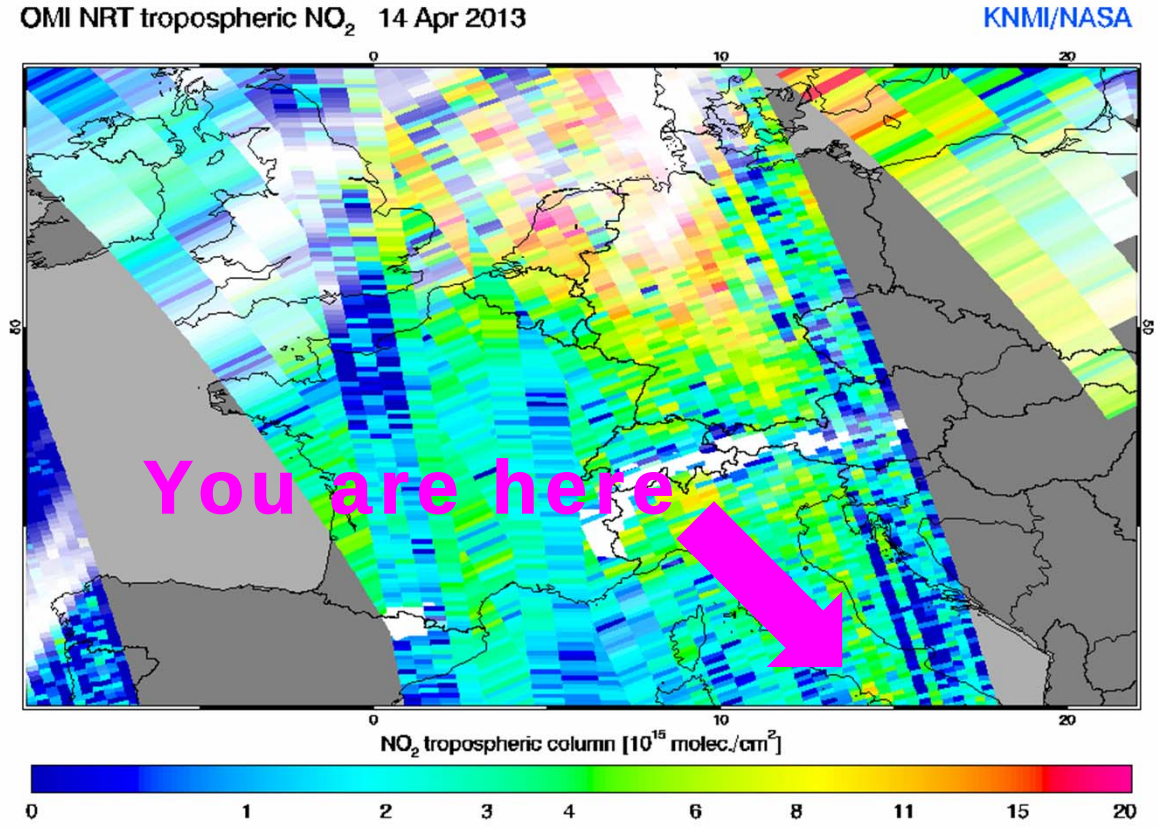
In the process of generating new products/tools/services, you need to:

- ***Understand*** *the documentation of all input data/models/etc.*

- ***Validate/inter-compare the input and output data*** *(as well tools/models)* ***used in each step of the product generation using best practise*** *approaches as outlined in the QA4EO framework*

- ***Document everything*** *and make it* ***readily available***

This appear very much in line with what EARSC is proposing with documentation



(MACC NRT O_3 and NO_2 service)

Thank you for your attention!

bojan.bojkov@esa.int