ECMWF Feature article

METEOROLOGY

Joint project trials new way to exploit satellite retrievals

Artist's impression of Jason-3 satellite (© EUMWTSAT)

Joint project trials new way to exploit satellite retrievals

Rossana Dragani

ECMWF is participating in the EU-funded AURORA (Advanced Ultraviolet Radiation and Ozone Retrieval for Applications) project, which is exploring new ways of exploiting the high-resolution data that will be provided with unprecedented accuracy by the Copernicus Sentinel-4 (S4), -5 (S5) and -5P (S5P) satellites.

AURORA will trial the use of data fusion of ozone retrievals from measurements in different spectral regions made by different sensors to reduce the amounts of data users need to handle, e.g. within data assimilation systems. Data fusion analytically combines atmospheric retrievals from different sources into a single, fused product characterised by greater quality than the individual retrievals. It will work with ad-hoc generated radiances that simulate measurements from the S4 and S5 satellites, which are scheduled to launch after the end of the project, and possibly with measurements from S5P, depending on its availability.

The project, which is funded under the European Commission's Horizon 2020 programme, started in February 2016 and will run for three years. ECMWF is making a major contribution to the project as one of the nine project partners (Box A). It also stands to benefit from it as AURORA offers an opportunity to refine the ozone assimilation while trialling it with fused ozone products. If successful, this approach could be exported to all applications that use ECMWF's Integrated Forecasting System, including the EU-funded, ECMWF-run Copernicus Atmosphere Monitoring Service (CAMS) and the Copernicus Climate Change Service (C3S).

The AURORA consortium   A

The consortium is led and coordinated by the N. Carrara Institute of Applied Physics, which is part of the Italian National Research Council, CNR-IFAC. It comprises nine partners, listed in the table, five of which are research organisations and four small and medium-sized enterprises (SMEs). The consortium also benefits from the support provided by the sub-contractor Resolvo s.r.l (Italy) for monitoring and communication. Airbus Defence and Space (Netherlands) acts as a third-party organisation, closely following progress on the opportunities explored by the project to exploit Copernicus data.

	AURORA partners		Country
Research Centres	CNR-IFAC	Consiglio Nazionale delle Ricerche – Istituto di Fisica Applicata Nello Carrara	Italy
	BIRA-IASB	Belgian Institute for Space Aeronomy	Belgium
	ECMWF	European Centre for Medium-Range Weather Forecasts	UK
	FMI	Finnish Meteorological Institute	Finland
	KNMI	Royal Netherlands Meteorological Institute	Netherlands
SME	DATACRAFT	DATACRAFT	Netherlands
	EPSILON	EPSILON International SA	Greece
	Flyby	Flyby SRL	Italy
	S&T	Science And Technology B.V.	Netherlands

Objectives

The project's objectives include:

- scientific objectives concerned with reducing the complexity of the high volume of Copernicus S4 and S5 data through a combination of data fusion and data assimilation

- the development of a technological platform for easy-to-use, efficient and quick data access

- the development of two operational, application-oriented services based on innovative mobile apps for UV dosimetry and tropospheric ozone monitoring.

B

Generating a satellite retrieval product

Satellite instruments do not directly measure the geophysical variables used in models (e.g. temperature or ozone). Instead, they measure the radiation at the top of the atmosphere at given frequencies. These measured radiances are related to the state of the Earth system, which is represented by the geophysical variables and hereafter simply referred to as 'the system', through the radiative transfer equation (RTE).

A radiative transfer model (RTM) aims to solve the RTE, i.e. to determine the state of the system given the measured radiances. This is referred to as solving the **inverse problem**. The process of determining the model equivalent of the measured radiances given the state of the system is referred to as solving the **forward problem**.

The retrieval process and its elements are schematically illustrated below, where panel a) represents the truth.

The **satellite measurements** (panel b) do not directly render an image of the truth, but something (the radiances) that can be transformed to represent it. Panel b) also symbolically shows that the sensitivity of a satellite instrument is not necessarily the same everywhere. Additionally, each remote instrument is designed to have a specific number of **degrees of freedom**. This is the maximum number of independent pieces of information the instrument is able to provide. In the picture, this could be the number of independent colours.
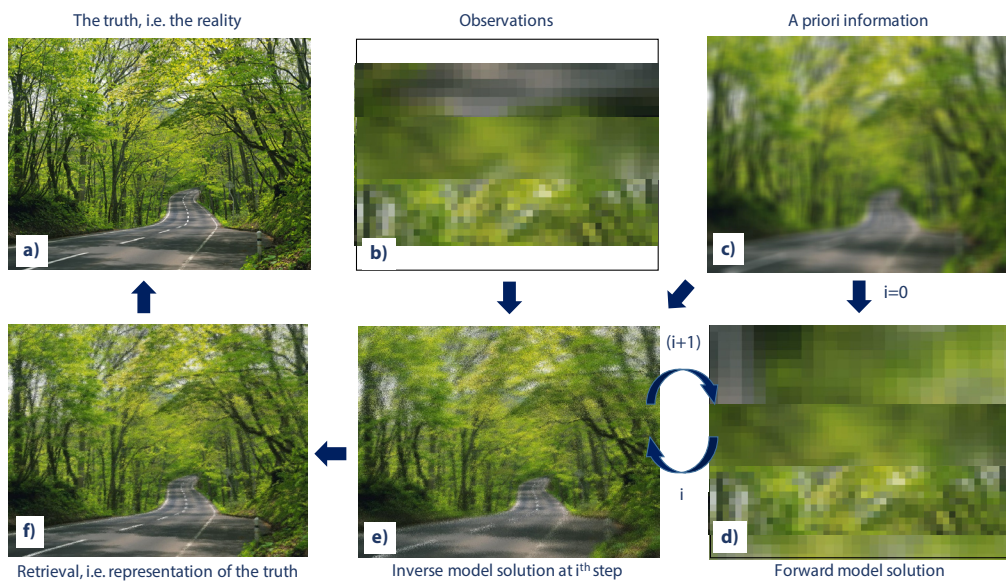
If we have a finite number of measurements, an infinite number of different solutions could produce the same measured radiances. This problem can be overcome by including prior knowledge of what the state of the system is. This **a priori information** (panel c) provides an indication of what the truth is, but the picture is smoothed and the details are difficult to see. The

closeness of the prior to the truth depends on its source. In many cases, a short-range forecast can be used as prior, and this can have a good degree of accuracy; in other cases, the retrieval algorithm can utilise a multi-year observation-based climatology, which can infer information on the main elements of the truth, i.e. a road and some trees, but not the specific details characterising the scene at the moment the 'true' picture was taken.

The observations, the prior, their error characterisation, and the forward model, which gives the model equivalent of the observations (panel d), are the elements used to derive a **model representation** of the truth (panel e). This model representation renders a picture that can largely be related to the truth, although many details are not well captured. This is not the final representation but it is used iteratively in the retrieval process together with the observations until convergence is reached.

The final **retrieval product** (panel f), also referred to as a Level 2 product to distinguish it from the radiances, which are normally called Level 1 data, is the best fit to the a priori information and the observations in a way that is consistent with their errors. The difference between the retrieval and truth is the so-called **retrieval error**. This error is a consequence of many factors; it depends on limitations and errors that affect the satellite measurements, and on limitations and errors in the RTM and the a priori information.

As for the observations, the retrieval quality is not the same everywhere but varies according to the instrument sensitivity, the quality of the prior, and the level of sophistication included in the forward model. The sensitivity of the retrieval product to the truth is referred to as the **Averaging Kernel Matrix**.

The truth, i.e. the reality   Observations   A priori information

Retrieval, i.e. representation of the truth   Inverse model solution at ith step   Forward model solution

Achieving these objectives, in particular the scientific ones, could make it easier for CAMS and C3S to exploit Sentinel data.

Core elements

The AURORA system revolves around two core elements: data fusion (DF) and the data assimilation system (DAS).

DF is an algorithm that analytically combines together atmospheric retrievals originating from different sources to produce a single, fused product characterised by greater quality than the individual retrievals. The fused product is fully characterised in terms of retrieval error, information gain, averaging kernels and number of degrees of freedom. A brief description of some of these concepts is provided in Box B.

Data assimilation is the process by which observations are incorporated into a numerical model and combined with prior knowledge in a way that is consistent with their uncertainties. It is also a way to blend different observations together, but unlike data fusion, it exploits the physical and dynamical coherence imposed by the laws of physics and ensures consistency between different physical properties.

DF makes it possible to substantially reduce the number of observations to pass to the DAS whilst representing a computationally more affordable alternative to producing simultaneous retrievals. The latter technique provides the best estimate of the observed atmospheric species because simultaneous retrievals take into account all the available information and rigorously handle non-linear effects. However, it can be difficult to implement because it requires a forward model that can simulate all the observations (made in different geometries and spectral regions). It is also computationally costly because the retrieval algorithm has to deal with a large amount of data. The implementation of DF is simpler but can lead to a loss of information, especially if the standard retrievals to be fused are available on different vertical grids. This possible loss of information is limited within the AURORA project by applying the Complete Data Fusion method (*Ceccherini et al.*, 2015). A brief introduction to the DF method used here is given in Box C.

In AURORA, DF is used on retrievals obtained from simulated measurements in different spectral regions from both the S4 and S5 platforms. The spectral ranges considered encompass UV, visible and thermal infrared (TIR) spectral bands. Total column ozone from measurements in the three spectral ranges and ozone profiles from the UV and TIR measurements are produced for both S4 and S5 observations. The complete DF is then applied to merge the standard ozone retrievals. The DF result is a synergistic ozone product fully characterised in terms of uncertainty, which is described by the variance-covariance matrix

---

Complete Data Fusion   C

The DF method used within the AURORA project is the Complete Data Fusion method presented by Ceccherini et al. (2015). It makes it possible to blend atmospheric vertical profiles retrieved from remote sensing measurements provided by different sources whilst limiting any loss of information. This is achieved by taking into account the retrieval errors of the retrieved profiles (i.e. the variance–covariance matrix), and the sensitivity of the retrieved profiles to the true profile (i.e. the averaging kernel matrix).

The DF method has been developed on the assumption that the forward models, which are used within the retrieval procedure to simulate the satellite measurements, are linear. On that assumption, it can be shown analytically that the solution obtained with complete data fusion coincides with the solution obtained with simultaneous retrieval. The DF algorithm further assumes that each retrieved product is available on a fixed vertical grid. Although not a strict requirement, ideally the standard retrievals obtained from different sources should be available on the same vertical grid.

The a priori information used in the individual retrieval procedures is removed before combining the information extracted from the measurements. The use of new a priori information is not strictly necessary, although it can be included in the DF to limit unexpected behaviours in the fused profile.

Because it is based on an analytical method, the DF does not require the use of any additional models, which can introduce external information. An important aspect of this technique is that the fused product represents the same physical variable as the original retrievals, making its use completely transparent to users.

The method can be used successfully to extend the vertical coverage of the final product, for instance, by exploiting complementary datasets with individual sensitivity limited in the vertical domain. However, it is unsuitable for spatial or temporal interpolation, or for extrapolation purposes. The application of DF can also result in a fused product of superior quality in altitude regions where both original measurements have information content different from zero.

(VCM), and vertical sensitivity of the retrieval to the true profile, which is described by the averaging kernel matrix (AKM). This fused ozone product is then exploited within a DAS to generate ozone analyses and forecasts up to about day 5.

Two DASs are available within the AURORA consortium: ECMWF's Integrated Forecasting System (IFS), and the Dutch national weather service's global chemical transport model (TM5). The assimilation of fused products has never been tested before and thus the impact on analysis and forecast performance compared with assimilating standard retrievals has never been assessed. Two experiments are envisaged with the two DASs, as schematically shown in Figure 1: one assimilating the fused product (labelled Exp 1) and the other assimilating standard ozone retrievals (labelled Exp 2). An additional baseline experiment that uses neither the fused product nor the standard retrievals is also envisaged. A thorough assessment of the resulting ozone analyses and forecasts will show whether and how well the DASs can exploit the additional information generated by DF.

The ozone analyses and forecasts generated by the two DASs are then used to calculate tropospheric ozone information and a UV index at the surface. These represent the AURORA products, which are planned to be used in two demonstration applications concerning air quality in major cities and personal UV dosimetry, respectively.
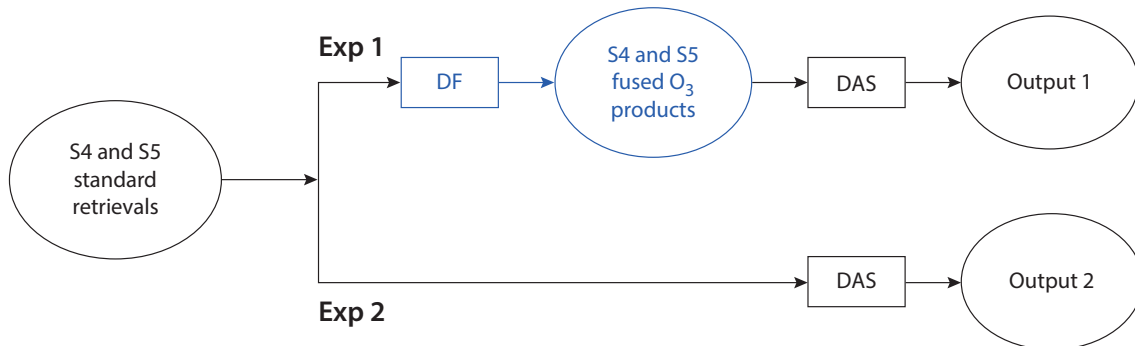
Figure 1 A schematic representation of the two experiments to be run with both DASs.

First big challenge

The AURORA project aims to demonstrate how user-driven applications can exploit the wealth of information provided by Copernicus Sentinel-4 (S4), -5 (S5) and -5P (S5P) measurements without the complication of handling huge data volumes. However, the S4 and S5 satellites are scheduled to be launched after the completion of the AURORA project in January 2019.

This poses a big challenge to the project, which needs to create a credible data flow and data infrastructure before the Sentinel datasets become available. For the purposes of the project, simulated observations which replicate operational Sentinel data as closely as possible need to be used. In practice, the simulated observations will be derived using the same instrumental characteristics as those of the Sentinel sensors.

A subset of NASA's second Modern-Era Retrospective analysis for Research and Applications (MERRA2) reanalysis is used to generate the atmospheric scenarios required to simulate the Sentinel measurements. The project does not have sufficient resources to perform a conventional Observing System Simulation Experiment (OSSE), which would require the simulation of all observations to be assimilated. However, the experiment's setup still meets the basic requirement for any OSSE study, which is to use two different models to simulate the measurements and to perform the assimilation. Based on this consideration, the list of required parameters, and the reanalysis characteristics (e.g. horizontal resolution), it was decided that the NASA MERRA2 reanalysis (*Bosilovich et al.*, 2015) was best suited to the needs of the AURORA project.

ECMWF's contribution

ECMWF has a crucial role within the AURORA project, with contributions covering three main work packages (WPs): WP3 (atmospheric scenarios and data simulation), WP4 (data fusion; data assimilation and forecasts; calculation of tropospheric ozone, calculation of the UV index at the surface; and development of the Prototype Data Processor), and WP8 (dissemination and exploitation).

In WP3, ECMWF is responsible for the preparation of the atmospheric scenarios that are used to simulate the Sentinel-4 and -5 observations. These consist of a selected set of model outputs retrieved from the MERRA2 reanalysis. This set of model outputs includes both meteorological fields and variables describing cloud and aerosol properties. This task has already been completed.

In WP4, ECMWF is both the WP leader and task contributor. In particular, it is responsible for running the IFS DAS to generate the global ozone analyses and forecasts that are then used to calculate the tropospheric ozone products and surface UV index. This contribution is mirrored at KNMI, where the TM5 DAS is also used to produce global analyses and forecasts of ozone. All partners involved in WP4 then contribute to the development and testing of the whole data processing chain. An overview of the role of all partners involved in WP4 is given in Figure 2.
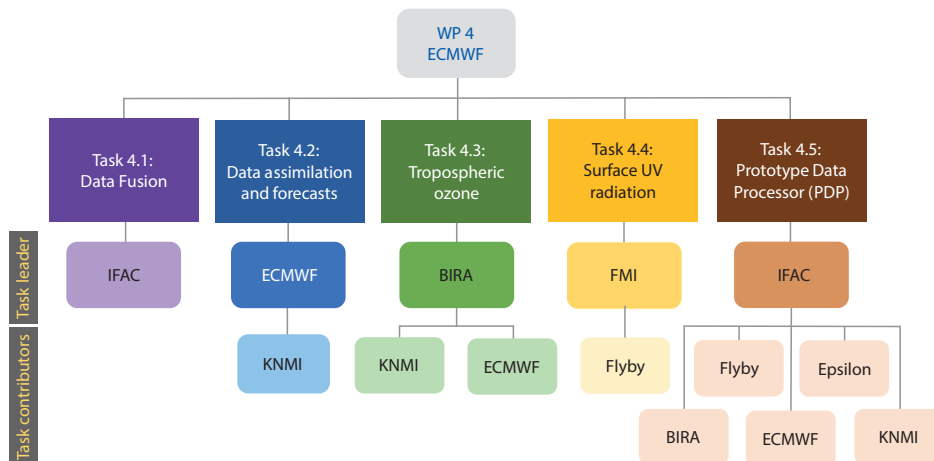

Figure 2 Roles of the AURORA partners involved in Work Package 4.

Figure 3 shows the data flow within WP4 and in relation to the work performed in other WPs. Conceptually, the data flow is represented by the black arrows. In practice, the Prototype Data Processor is being developed to make data accessible from a shared geospatial database through user-customisable dashboards. This is represented by the blue arrows.

In WP8, ECMWF contributes to the outreach activities of the consortium, in particular by promoting the AURORA work and outcomes to the Copernicus Climate Change and Atmosphere Monitoring Services.
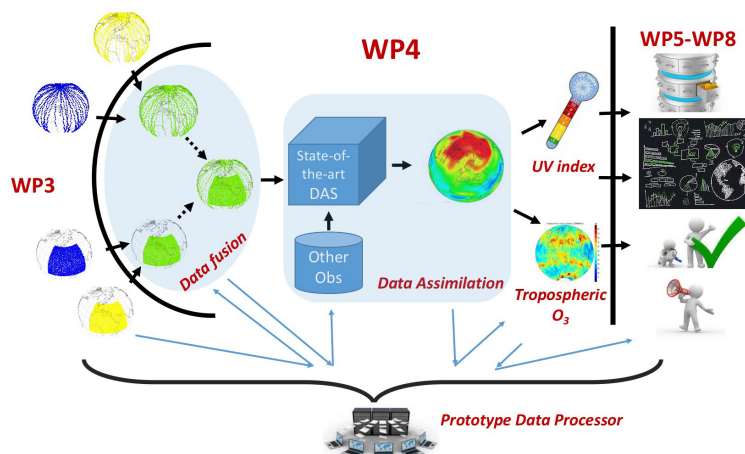

Figure 3 Envisaged data flow within the AURORA project (black arrows). In practice, all data pass through the Prototype Data Processor (blue arrows). ECMWF leads WP4, which includes five main tasks: data fusion; data assimilation; the calculation of tropospheric ozone; the calculation of a UV index; and the development of the Prototype Data Processor. Work package 3 focuses on the Sentinel data simulation and retrieval; work packages 5 to 8 cover the following aspects: data acquisition and storage (WP5), web services and data visualisation (WP6), data validation and quality assurance (WP7), and outreach and dissemination (WP8).

Expected impact

The AURORA project is expected to have a significant impact on both the scientific and the technological front. It will also develop applications that could be useful to a range of users, including the general public.

On the scientific front, the application of data fusion methods is completely new, especially in combination with data assimilation. This has generated considerable interest, especially since the fused products are of greater quality than individual retrievals while still representing the same physical variable as the original, unfused data.

Figure 4 Representatives of the AURORA consortium and members of the External Expert Advisory Board at the first AURORA progress meeting in Reading, UK, on 20 and 21 July 2016.

The wealth of data that the Copernicus S4 and S5 will deliver (an estimated 27.7 million measurements per day at solar zenith angles smaller than 80°) is overwhelming and perhaps even prohibitive for many key players in the scientific community, industry, and the public sector. If the AURORA project succeeds in simplifying access to data and its information content, and even in providing added-value products, then new possibilities can open up, especially in terms of applications and services. Potential users outside the AURORA consortium will be welcome to test the AURORA datasets in demonstration applications once the AURORA products become available towards the end of the project. However, it is worth remembering that simulated observations will be used instead of real S4 and S5 measurements. Thus, caution will need to be exercised when drawing conclusions from using AURORA data.

Two technological elements will be developed: the AURORA interface and a web-based Geographic Information System platform providing automatic access to harmonised data and to a user-friendly customised interface. The latter will provide advanced techniques for data visualisation. All these elements are expected to significantly facilitate the use of Copernicus Sentinel data by a wide community of scientists and application developers. They will also suggest a possible model for operational data dissemination to users.

Further reading

Bosilovich, **M. G.**, **S. Akella**, **L. Coy**, **R. Cullather**, **C. Draper**, **R. Gelaro**, **R. Kovach**, **Q. Liu**, **A. Molod**, **P. Norris**, **K. Wargan**, **W. Chao**, **R. Reichle**, **L. Takacs**, **Y. Vikhliaev**, **S. Bloom**, **A. Collow**, **S. Firth**, **G. Labow**, **G. Partyka**, **S. Pawson**, **O. Reale**, **S. D. Schubert**, & **M. Suarez**, 2015: MERRA-2: Initial Evaluation of the Climate, *NASA TM-2015-04606*, **43**, available from http://gmao.gsfc.nasa.gov/pubs/tm/.

Ceccherini, **S.**, **U. Cortesi**, **S. Del Bianco**, **P. Raspollini**, & **B. Carli**, 2010: IASI-METOP and MIPAS-ENVISAT data fusion, *Atmos. Chem. Phys.*, **10**, 4689–4698.

Ceccherini, **S.**, **B. Carli** & **P. Raspollini**, 2015: Equivalence of data fusion and simultaneous retrieval, *Optics Express*, **23**, 8476–8488.

Dragani, **R.**, 2016: A comparative analysis of UV nadir-backscatter and infrared limb-emission ozone data assimilation, *Atmos. Chem. Phys.*, **16**, 8539–8557, doi:10.5194/acp-16-8539-2016.