ENVRIplus DERIVERABLE

D2.3

Harmonization strategy report toward support to space mission and assimilation system based on RI data

WORK PACKAGE 2 – Metrology, quality and harmonization

LEADING BENEFICIARY: **Italian National research Council (CNR)**

Author(s):	Beneficiary/Institution
Simone Lolli	Consiglio Nazionale delle ricerche (CNR)
Lucia Mona, Gelsomina Pappalardo	Consiglio Nazionale delle ricerche (CNR)
Andreas Petzold	Juelich University
Jean-Daniel Paris	ICOS-RI
Glenn Nolan	EUROGOOS
Vito Vitale	Consiglio Nazionale delle ricerche (CNR)

Accepted by: **Vito Vitale, WP 2 leader**

Deliverable type: REPORT

Dissemination level: PUBLIC

Deliverable due date: 31.03.2019

Actual Date of Submission: 28.07.2019

Table of Contents

Executive Summary

Deliverable 2.3 main purpose is to describe the state-of-the art of the current Research Infrastructures (RI), both in the atmospheric and marine domain, especially for validation of the satellite measurements and for data assimilation into numerical forecast models. Being the atmospheric domain RIs independently established, often on a federated basis, with independent strategies (between the different RIs), a lot of efforts are currently put to standardize RIs measurements into compliant products to meet the requirements of data satellite validation and for model assimilation. The immediate goal of each atmospheric RI is to provide standardized products and heterogeneous time series directly comparable without introducing biases in the scientific results, i.e. ACTRIS and ICOS RI's are currently putting lot of efforts in providing harmonized and scientifically significative in metrological sense trace gases concentration measurements that will be available for satellite validation.

With respect to the atmospheric domain, the marine Research Infrastructures have a different outlook, as EUROGOOS is coordinating each different marine Research Infrastructure to provide harmonized scientific data, in the frame of Copernicus Services' use of meteorological, climatological, and hydrological data for Emergency Management System (CEMS). In this framework EUROGOOS is in charge to harmonize the different *in-situ* marine RIs observations (described into this deliverable) to establish a single interface access to be implemented into CEMS.

The principal result put in evidence by this deliverable is that the benefits of integrating ground-based with satellite data is bidirectional, as satellite data contributes to improve the information at surface and *vice-versa*. For this reason, future strategies should be implemented to establish tight synergies among satellite and ground-based observations

ABSTRACT

The European atmosphere and marine research infrastructures monitor the atmosphere and oceans fundamental variables, through *in-situ* and remote sensing measurements, linking global and regional scales at different temporal resolutions. The measured variables are fundamental indicators to quantitatively assess Earth system health. The ultimate goal of the atmospheric and marine research infrastructures is to provide a high-quality measurement database to support both the space and in-situ measurement components of the European Union COPERNICUS program, especially for the Sentinel missions.

With respect to the atmospheric domain, different processes are increasingly the focus of many societal and environmental challenges, such as air quality, health, sustainability and climate change, i.e. reliable predictions of the future climate using climate models are central and fundamental requirements for determining future mitigation strategies. In this framework, Research infrastructures such as ACTRIS (Aerosols, Clouds and Trace gases Research Infrastructure) provides a platform for researchers to combine their efforts more effectively making available high-quality observational data of aerosols, clouds optical and microphysical properties and trace gases concentrations openly available to anyone who might want to use them. IAGOS (In-service Aircraft for a Global Observing System) provides instead a database of airliner measurements for users in science and policy while the Integrated Carbon Observation System (ICOS) provides data on greenhouse gas concentrations. ICOS, as part of the European environmental Research Infrastructure landscape, contributes at different levels both to assimilate and validate the Copernicus Atmosphere Monitoring Service (CAMS), one of the six services part of Copernicus Earth Observation program.

With respect to the aquatic domain, research infrastructures such as EuroGOOS contribute to the operation of COPERNICUS Marine Environment Monitoring Service (CMEMS) through a broad range of activities, which include identifying priorities, enhancing cooperation and promoting the benefits of operational oceanography to ensure sustained observations made in Europe's seas underpinning a suite of fit-for-purpose products and services for marine and maritime end-users. In the same framework, Euro Argo ERIC is now the single most important in-situ observing system for operational oceanography. Its main focus is maintaining the global array of measurements that it is essential for the long-term sustainability and evolution of the CMEMS. Similarly, the objective of JERICO-NEXT is the strengthening and enlarging of a solid and transparent European network of operational services for the timely, continuous and sustainable delivery of high quality environmental data and information products related to marine environment in European coastal seas.

Project internal reviewer(s):

Project internal reviewer(s):	**Beneficiary/Institution**
Helen Glaves	British Geological Surveys
Paolo Laj	Université Grenoble-Alpes

Document History:

Data	**Version**
30-05-2019	First draft for comments
30-07-2019	Final Version

1 The Atmospheric Case

In-situ and remote sensing Research Infrastructures (RI) measurements are crucial to validate and improve, i.e. through the assimilation, the current numerical weather prediction model forecasts and the Copernicus Atmosphere Monitoring Service (CAMS), one of the six services part of the Copernicus Earth Observation program. CAMS is implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission with the support of 34 states. ECMWF produces and disseminates numerical weather predictions to its member states. CAMS combines the expertise and infrastructure of European countries to provide a wide range of unique services. To collect all observations needed to produce CAMS services, ECMWF collaborates with European Space Agency (ESA), the European Organization for the exploitation of Meteorological satellites (EUMETSAT) and other organizations providing satellite and in-situ observations. The atmospheric RI observations are related to the satellite observation validation, especially the Sentinel and the ADM-Aeolus missions. The RIs directly involved in the atmospheric case are:

- ACTRIS (Aerosols, Clouds and Trace gases Research Infrastructure), a pan-European initiative consolidating actions amongst European partners producing high-quality observations of aerosol and cloud optical and geometrical properties and trace gas concentrations both with *near-surface* and remote sensing systems accompanied by ancillary measurements of meteorological and radiation quantities. Detailed info on Actris RI can be found in section 1.1

- IAGOS (In-service Aircraft for a Global Observing System) is a distributed European Research Infrastructure with Members from Germany, France and the UK, organized as international non-profit association AISBL with its seat in Brussels, Belgium. IAGOS operates a global-scale monitoring system for atmospheric chemical composition by using a fleet of commercial passenger aircraft equipped with miniaturized and fully automated instrumentation. Detailed information on ICOS RI can be found in section 1.2

- ICOS (Integrated Carbon Observation System) is a pan-European research infrastructure founded in 2008, with the Head Office located in Helsinki, Finland. It provides data on greenhouse gas concentrations and is thus part of the European environmental Research Infrastructure landscape. Currently, ICOS Research Infrastructure has more than 100 stations in 12 European countries. Detailed information on ICOS RI can be found in section 1.3

- SIOS (Svalbard Integrated Arctic Earth Observing System) is a distributed international research infrastructure for Arctic Earth System Science, coordinating a regional observing system for long-term measurements in and around Svalbard. SIOS became a Copernicus Relay in 2016. Detailed information on SIOS RI can be found in section 1.4

Table 1 summarizes the activities of all the atmospheric domain RIs, specifying the measurement type, coverage and key-relevant parameters.

Research Infrastructure	Measurement Type	Coverage	Domain	Key relevant parameters	Website	Data Policy
ACTRIS	*In-situ* Remote Sensing Laboratory	100 stations mostly in EU	From surface to stratosphere	Remote:clouds, aerosols, trace gases In-Situ:cloud, aerosols, trace gases Lab:cloud and aerosols	https://www.actris.eu	Open data (owned by ACTRIS partner)
IAGOS	*In-situ*	Global	Flight Level and vertical profiles	In-situ: clouds, aerosols, trace gases	http://www.iagos.org	Open data policy
ICOS	*In-situ*	130 stations in EU	Surface-150m	In-situ:Greenhouse gases On board of vessels	https://www.icos-ri.eu	Free upon giving appropriate credit
SIOS	*In-situ* Remote sensing Lab	Svalbard Island		On demand	https://sios-svalbard.org	Depending on particular situation

Table 1 Summary of the atmospheric Research infrastructures main characteristics

1.1 ACTRIS – Aerosol, Cloud and Trace-gases Research InfraStructure

Contact: Lucia Mona, lucia.mona@imaa.cnr.it, gelsomina.pappalardo@imaa.cnr.it

1.1.1 Overview

The Aerosols, Clouds and Trace gases Research Infrastructure (ACTRIS) is a distributed infrastructure dedicated to high-quality observation of aerosols, clouds, trace gases and exploration of all their interactions. It will deliver high-precision data, services and procedures regarding the 4D variability of clouds, short-lived atmospheric species and the physical, optical and chemical properties of aerosols to improve the current capacity of analyzing, understand and predict past, current and future evolution of the atmospheric environment. ACTRIS serves a vast community of users working on observations, experiments, models, satellite data, analysis and predicting systems and offers access to advanced technological platforms for exploration of the relevant atmospheric processes in the fields of climate change and air quality

ACTRIS is the pan-European Research Infrastructure (RI) composed of 9 connected elements: distributed National Facilities (observation platforms and exploratory platforms) both in Europe and globally, and 8 Central Facilities (Head Office, Data Centre and 6 Topical Centers). The 6 Topical centers are the aerosol, cloud and trace gases remote sensing topical centers, and the corresponding 3 ones for the in situ measurements. ACTRIS provides access to its facilities, open-access data, measurement support, instrument calibration and development, and training to various user groups. ACTRIS entered in the ESFRI roadmap in 2016 as a new project and it is expected to be fully operational in 2025.
All different atmospheric predictions use complex models that are underpinned by observations. Without high quality observation data to constrain predictive models, the forecast of any atmospheric variable is highly unreliable. ACTRIS focuses on producing high-quality observations of short-lived climate forcers (SLCFs) and other short-lived atmospheric components. These components have a residence time in the atmosphere ranging from hours to few weeks, which differentiates them from long-lived greenhouse gases. The short lifetimes make their concentrations highly variable in time and space and involve processes occurring on very short timescales. These considerations separate SLCFs from long-lived greenhouse gases (LLGGs), and calls for a distributed observatory (WMO, 2012). Such an observatory is provided by ACTRIS consisting of different stations in Europe and outside Europe (see Figure 1.1.1), and a number of Central

facilities fundamental for the provision of harmonized high-precision data required by the scientific community. The services provided by ACTRIS are fundamental to address the following scientific questions:

• To quantitatively assess how aerosols and trace gases affect Earth's radiation balance.

• To assess the cloud feedback to the total Earth radiative budget. There are large uncertainties due to the complexity of cloud systems and how their formation and lifetime is influenced by aerosol concentration.

• To reduce air pollution and related adverse effects on health and ecosystems. It is well established that aerosol particles, at concentrations typically found across Europe, give rise to severe and unacceptable health effects in the European population (WHO, 2013). The situation is even exacerbated in other world regions.

• To fill the major gaps in knowledge quantifying the impact of climate-induced feedback mechanisms on atmospheric composition. The number of change drivers is very large with strongly coupled systems. An additional level of complexity is linked to the issue of anthropogenic-induced climate–chemistry interaction.

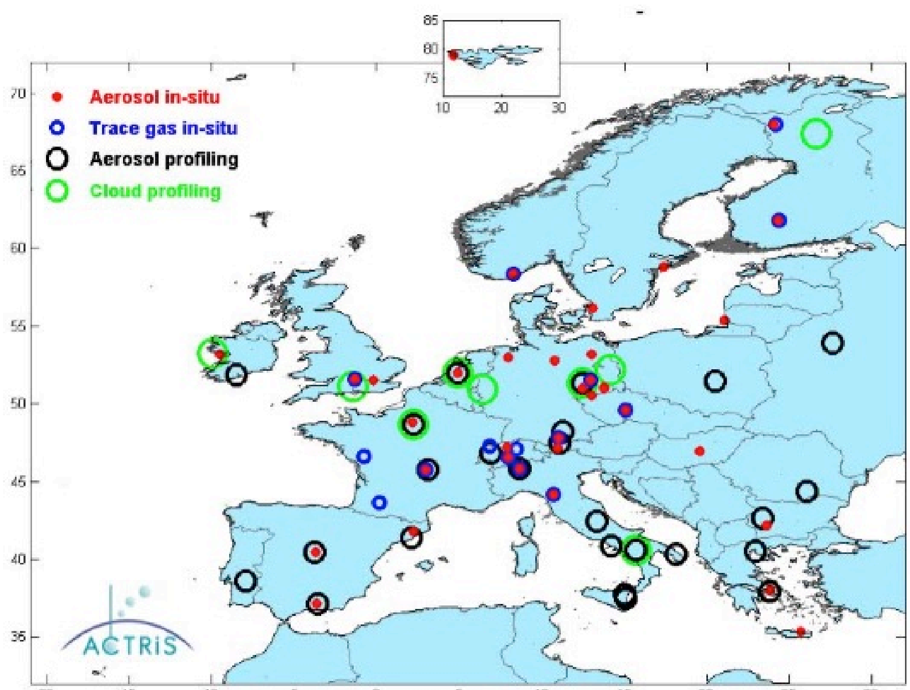


Figure 1.1.1 ACTRIS Current stations

1.1.2 ACTRIS for validation and assimilation purposes

Since early 2000, different components of ACTRIS (acting as research networks at that time) were involved in satellite missions or model evaluation related activities. Typically, those activities were inter-comparing ground-based measurements with the corresponding satellite mission, e.g. passive ground-based aerosol vs.

passive satellite aerosol measurements, or for very specific research interests (i.e. volcanic plume dispersion, large forest fires, dust outbreaks, haze events). Such activities were fairly coordinated at network level, without standardized procedures and protocols, but *ad-hoc* solutions were adopted for the specific applications.

1.1.2.1 Satellite validation and exploitation

Since the beginning, all the ACTRIS components worked on satellite data validation as one of the main research network activities. Recent ACTRIS related activities permitted to validate satellite observations and increase knowledge about the atmospheric content. In particularly, the lidar network component of ACTRIS, i.e. the European Lidar Network EARLINET, main objective is to quantitatively assess the atmospheric profile of the optical, microphysical and geometrical properties of aerosol and clouds through quality measurements obtained from a network of multi-wavelength Raman lidar prototypes. In this framework, EARLINET database network lidar measurements are crucial to validate satellite missions as CALIPSO, the first satellite mission with a lidar onboard. The mission is specifically designed to study the vertical profile of the aerosol and cloud optical and geometrical properties. ESA funded a EARLINET/ACTRIS project for CALIPSO data exploitation and for paving the way of next ESA missions with lidar onboard. During this study, an observational strategy at network level was designed and the comparison at the raw data level showed the absence of significant biases in the signal (Pappalardo et al., 2010) with CALIPSO. The comparisons, in terms of level 2 products, showed the new CALIPSO features but also made evident some issues in properly layering and typing the atmospheric aerosols (Pappalardo et al.,2010, Wandinger et al., 2010). These deficiencies were further investigated in the analysis of CALIPSO climatological products, and some solutions were proposed to address the problem, i.e. CALIPSO needed assumptions taken from the EARLINET/ ACTRIS (Papagiannopoulos et al., 2016; Figure 1.1.2) database. Provided suggestions were considered by the CALIPSO team for the new released version of CALIPSO data, it gave a better insight into the aerosol condition on a global scale.

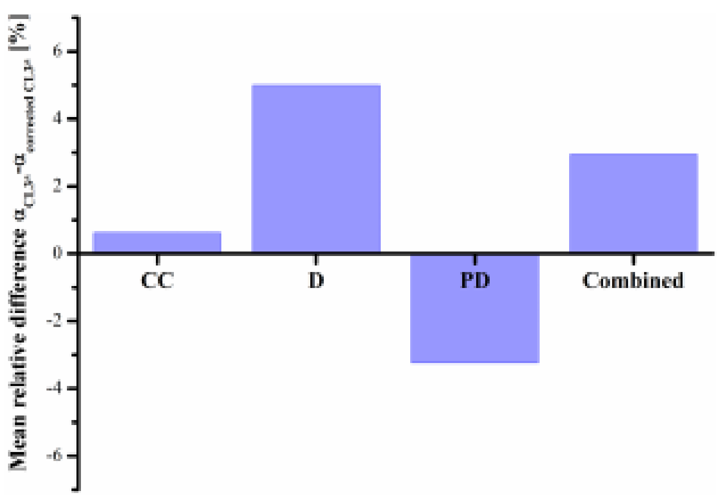


Figure 1.1.2 Estimation of the mean difference between the CALIPSO version 3 aerosol extinction value and the one obtained using the EARLINET/ACTRIS lidar ratio values into the CALIPSO algorithm for 4 types of aerosol (CC=clean continental; D=dust, PD= polluted)

The combination of advanced knowledge provided by EARLINET/ACTRIS on aerosol typing together with the geographical coverage of CALIPSO datasets brought to LIdar climatology of Vertical Aerosol Structure for space-based lidar simulation studies (LIVAS; Amiridis et al., 2015), which is a powerful tool for lidar end-to-end simulations of realistic atmospheric scenarios, climatological investigation and long term model evaluation. The representativeness of aerosol profiles for long range transport events were studied in a CALIPSO-EARLINET integrated study through the collection of ground-based and satellite measurements at different temporal and spatial resolutions. Additional validation efforts were based on a specific case study analysis, i.e. the big and highly-impacting volcanic eruption of Eyjafjallajökull in 2010. Volcanic ash and aerosol height retrievals by GOME-2 and IASI were validated through the extensive measurements collected during the biggest aviation crisis ever observed in Europe (Balis et al., 2016). Remote sensing of aerosol and cloud remote optical and microphysical properties are two important ACTRIS activities. Observation from those activities are currently involved in a synergic way to validate aerosol and clouds product retrievals from recently launched ESA sensors as Aeolus lidar and Sentinel5P. Another ACTRIS component addresses trace-gas concentration measurements. Four decades ago, few ozone monitoring stations were already used as ground-based references for the geophysical validation of Total Ozone Mapping Spectrometer (TOMS) column and SAGE-II and SBUV/2 profile data. These pioneering validation activities were progressively developed to encompass all types of Network for the Detection of Atmospheric Composition Change (NDACC) instruments and their complete portfolio of species and parameters. Validations, based on single instruments at single stations, were expanded to more comprehensive assessments using the network as a whole. NDACC data have also been used to assess the stability and mutual consistency of multiple satellite data records across a multi-decadal period. A comprehensive review of the validation activities performed and the associated results is reported in De Maziere et al., 2018. Here an example is reported in Figure 1.1.2, showing the success of the Montreal protocol in reversing the ozone depletion tendency observed until 2000 at all sites.

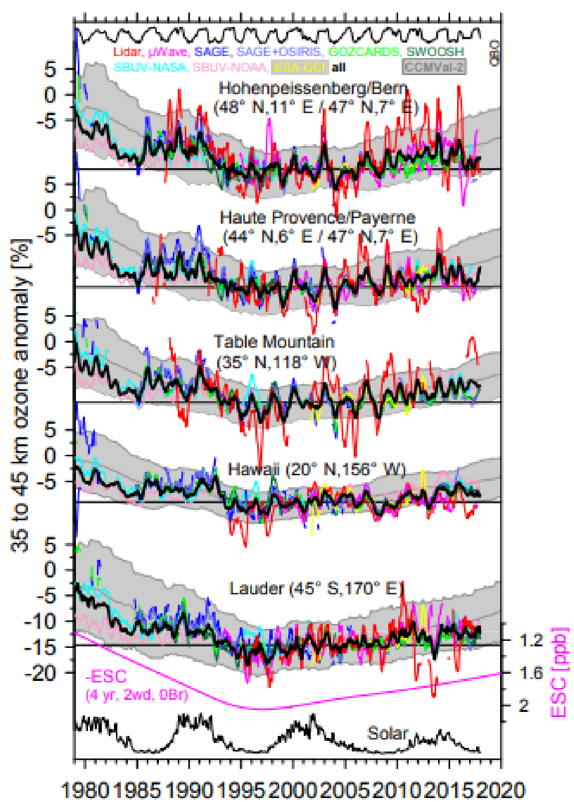


Figure 1.1.3 upper-stratospheric ozone time series from satellite data (SAGE (Stratospheric Aerosol and Gas Experiment), OSIRIS (Optical Spectrograph and InfraRed Imager System), ESA-CCI), SWOOSH (The Stratospheric Water and OzOne Satellite Homogenized and GOZCAR

1.1.3 ACTRIS data for Model evaluation and assimilation

The cooperation between ACTRIS community and the modelers was initiated well before the integration of the EUropean Supersites for Atmospheric Aerosol Research (EUSAAR), EARLINET, Cloudnet and NDACC in a unique comprehensive research infrastructure. Since the beginning, cloud profile observations were used for evaluating and scoring numerical weather forecasts (see e.g. Illingworth et al., 2007). Cloudnet/ACTRIS reports already show the datasets resulting from such comparisons in terms of climatologies (means, distribution, skill scores). In addition, extended skill score comparison for the different models is available in terms of long-term (inter-annual variability), forecast lead time, model version, model length-scale and location). In particular, the SEDI score (Symmetric extremal dependence index) is investigated to assess the general performance of a forecast system and its ability to predict extreme (rare) events.

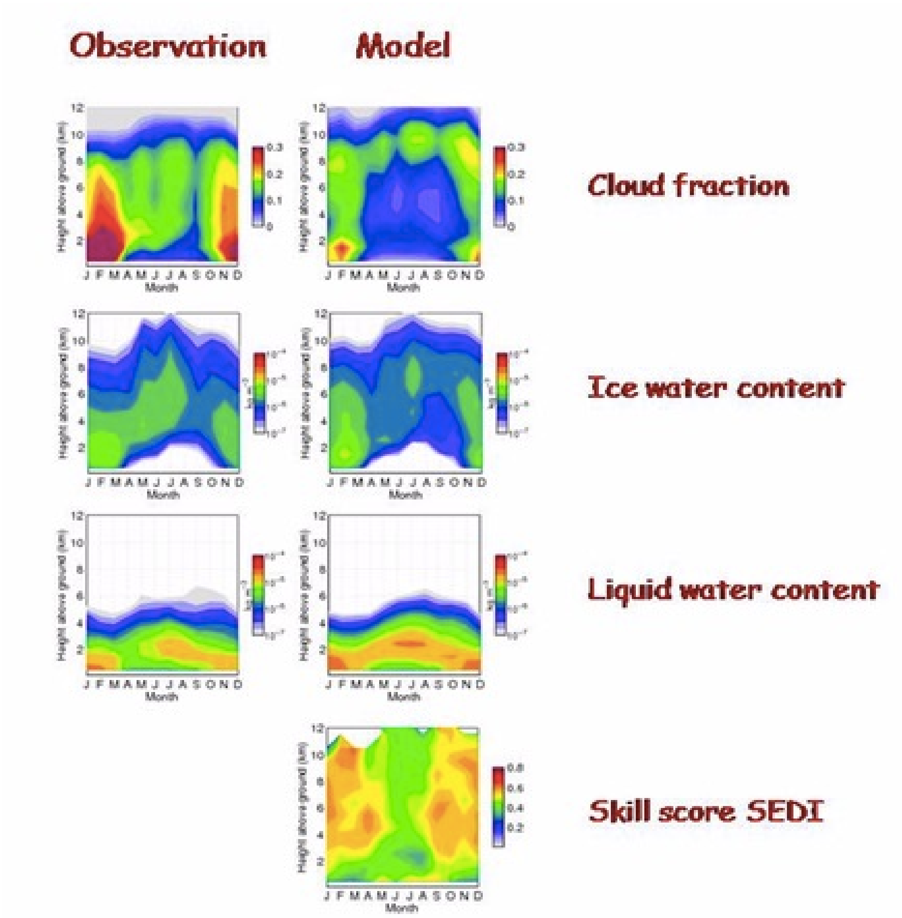


Figure 1.1.4 comparison between Cloudnet measurements and models about Cloud fraction, ice and liquid water content, and Cloud fraction SEDI skill score, this last showing the low capability of the model to reconstruct the extreme observed values in winter time cloud fraction.

A multi-decadal cooperation established between the ground-based measurement community and AEROCOM project, led to a long-term comparison and trend analysis as reported in the EMEP reports (e.g. http://emep.int/publ/reports/2018/EMEP_Status_Report_1_2018.pdf). During the ACTRIS2 project, a tool for the visualization of trends as modeled by AEROCOM and as observed by surface, Aeronet and EARLINET/ ACTRIS components has been developed. Recently, a prototype for evaluation and online verification of models became available from the CAMS site using aerosol scattering coefficients from numerous sites within ACTRIS and GAW (Figure 1.1.4)

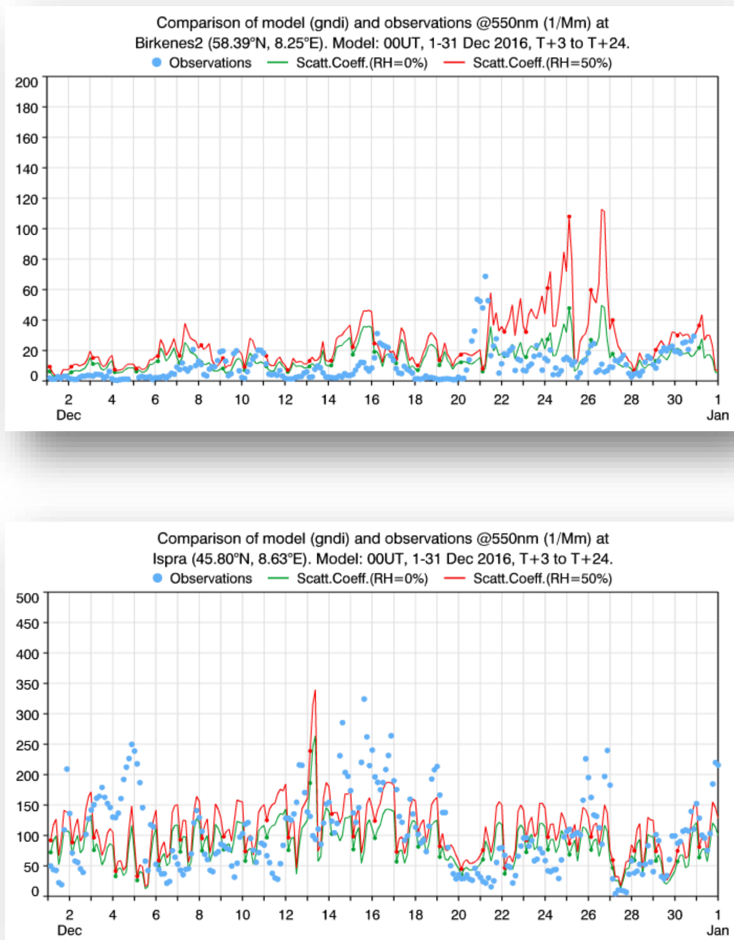


Figure 1.1.5 Example of time series of CAMS modeled and ACTRIS observed values for the scattering coefficient at 550 nm. Two sites with very different aerosol content are reported. A model overestimation is found at remote and clean site of Birkenes (top panel) while a good agreement in terms of values and temporal behavior is observed for the polluted site of Ispra (bottom panel).

Similarly EARLINET/ACTRIS profiles are used as references to evaluate the ground aerosol concentration for operational validation of the regional ensemble model for Copernicus ENSEMBLE air quality forecast (ENS-2016 and ENS-2017). The example in Figure 1.1.5 shows how much the model underestimates the surface concentrations at Leipzig station, along with too much mixing in the upper levels. Model evaluation activities were also carried out using EARLINET/ACTRIS observations for some specific case studies like in the famous case of the Eyjafjallajökull eruption in 2010 (Matthias et al., 2012). The cooperation with dust forecast modelers started in early 2000 for the aerosol vertical profiling communities, when an alerting system was set up for triggering measurements at the station for desert dust investigation. This cooperation has lead to the development of suitable methodologies for the evaluation of model capability of reproducing dust vertical profiles (Mona et al. ACP 2012, Binietoglou et al., 2013). Additionally long-term desert dust aerosol vertical profiles acquired since 2000 by EARLINET/ACTRIS will be used for the evaluation of the BSC NMMB-MONARCH reanalysis under the DustClim- ERA4CS project.

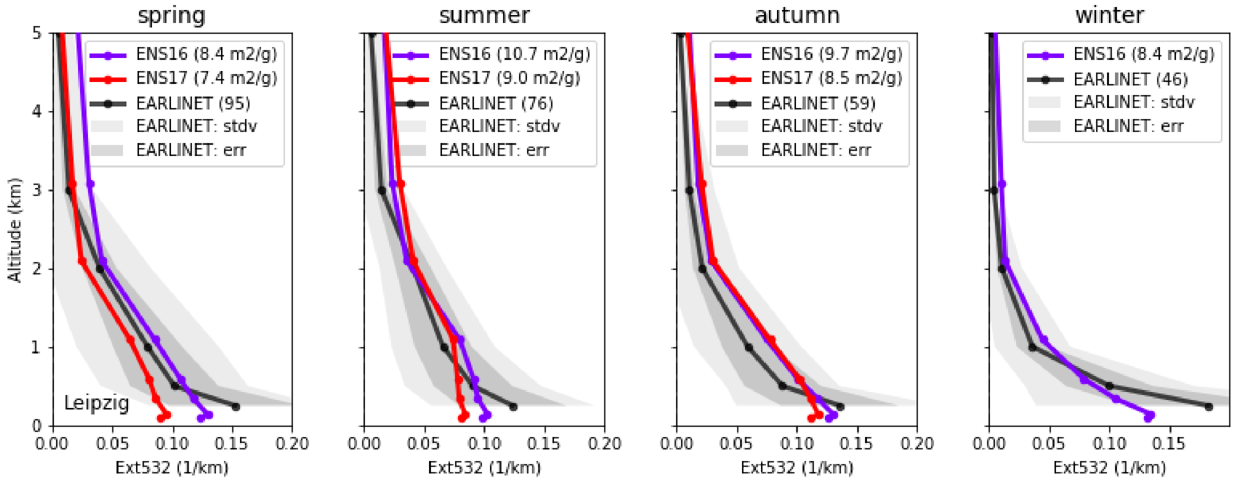


Figure 1.1.6 Comparison of aerosol extinction profiles in terms of climatological seasonal averages. The ensemble means for 2016 and 2017 are reported together with EARLINET averaged.

Remote sensing observations of trace gases are also used as independent reference data within CAMS and C3S. In CAMS-84, ACTRIS provides data for O3, NO2, HCHO, aerosol, CO, CH4 validation. The multi-decadal available measurements made ACTRIS trace gases remote sensing observations relevant for C3S-LOT311a service, where it is in charge for the provision of climate data records (long-term homogenized time series) of O3, CO, CO2, CH4.

Assimilation of aerosol vertical profiles is a nowadays a cutting edge topic. Some pilot exercises have been recently performed or were performed within the EU Horizon2020 projects "ACTRIS 2" and "EUNADICS – AV". Some assimilation exercises are currently in progress for ensemble model developed in EUNADICS_AV and for ECMWF in ACTRIS 2 projects for the Eyjafjallajökull volcanic case. Additionally, desert dust backscatter profiles acquired during the ACTRIS Summer 2012 measurements campaign have been assimilated in the BSC dust model simulations from the chemical weather prediction system NMMB-MONARCH. The assimilation is run with a 1 hour time resolution for the analysis calculation within a *24* hour assimilation window. The assimilation of measured lidar profiles helps in correcting inconsistencies between observed and simulated dust plumes.

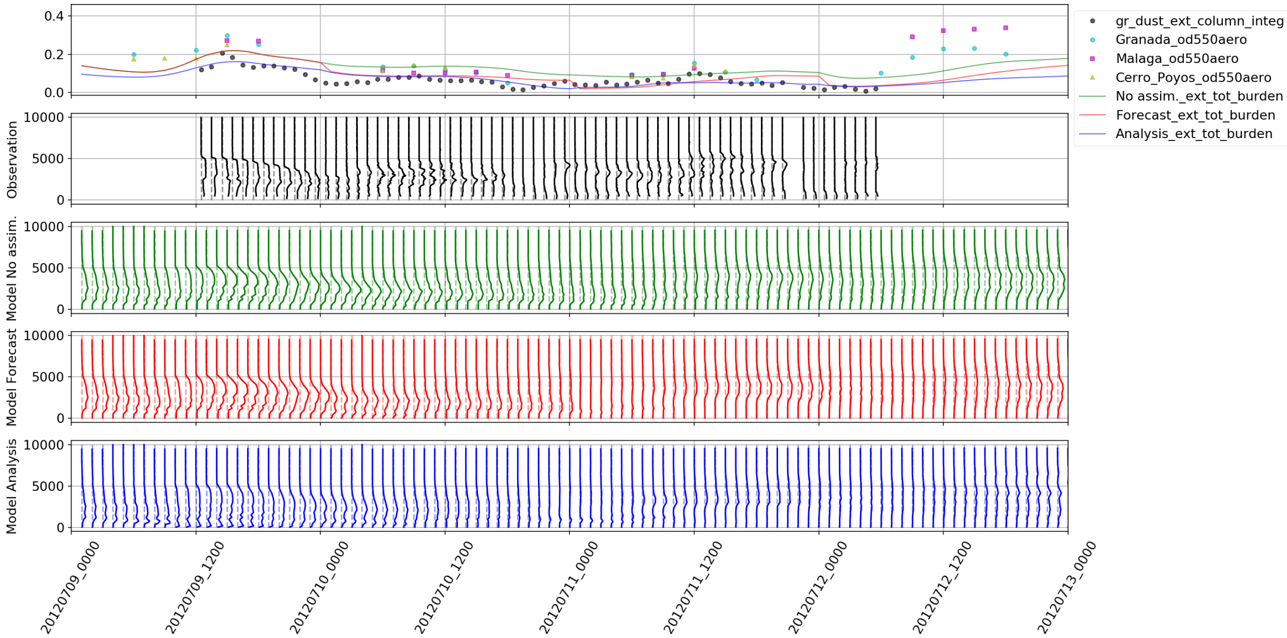


*Figure 1.1.7 **:** Aerosol optical depth (AOD) at 550nm from three AERONET stations (Granada, Malaga and Cerro Poyos), column integrated extinction of lidar dust profiles (lidar equivalent AOD, black), extinction total burden of analysis (analysis equivalent AOD, blue), analysis-initialized forecast (AOD, red), forecast with no assimilation (AOD, green) at 532nm for the experiment period, from 9 July 2012 to 12 July 2012 (top panel). Lower panels contain dust extinction profiles from assimilated lidar observations (black), from a simulation without data assimilation (green), from model forecast initialized from an analyses (or first-guess; red), from model analysis (blue). The vertical coordinate for the latter is altitude (m) above the surface, and the horizontal scale is the same for all the profiles in this figure.*

1.1.4 Future plans

ACTRIS measurements and data user interactions fostered innovation and developments inside the ACTRIS community (gathering information about user requirements and acquiring experience in user driven approaches) and outside ACTRIS community (better insight of the atmospheric status and new features and progresses in satellite/model data). This interaction was part of the development progress which promotes ACTRIS from a pure research project to a consolidated Research infrastructure level. Links with users (scientific, SMEs, society, citizens) drives the development and dissemination of the integration tools to fully exploit the use of multiple atmospheric techniques at ground-based stations. Moreover, centralized and controlled procedures of the climatological ACTRIS products can be relevant for the reevaluation of analyses, satellite measurement cross-validation, and bridging between different satellite datasets. In this context ACTRIS is providing level 3 (climatological) products for cloud remote sensing observations and for aerosol remote sensing (end of March 2019). Open data policies, and data FAIRness and traceability aspects as designed for the ACTRIS Data center will allow broad evaluation of Copernicus model(s), satellite missions and models.

1.1.5 Acronyms

ACTRIS Aerosols, Clouds and Trace Gases
AEROCOM: Aerosol Comparisons between Observations and Models
AERONET: Aerosol Robotic Network
AOD: Aerosol Optical Depth
BSC: Barcelona Supercomputing Center

C3S: Copernicus Climate Change Service
CALIPSO: Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CAMS: Copernicus Atmosphere Monitoring Service
EARLINET: European Aerosol Research Lidar Network
ECWMF: European Centre for Medium-Range Weather Forecast
EMEP: European Monitoring and Evaluation Programme
ERA4CS: European Research Area for Climate Services
ESFRI: European Roadmap for research infrastructures
EUNADICS-AV: European Natural Airborne Disaster Information and Coordination System for Aviation
EUSAAR : European Supersites for Atmospheric Aerosol Research
GAW: Global Aerosol Watch
GOME: Global Ozone Monitoring Experiment
IASI: Infrared Atmospheric Sounding Interferometer
LIVAS: LIdar climatology of Vertical Aerosol Structure for space-based lidar simulation studies
MONARCH Multiscale Online Nonhydrostatic AtmospheRe CHemistry mode
NDACC: Network for the Detection of Atmospheric Composition Change
FAIR: Findable, Accessible, Interoperable and Reusable
SAGE: Stratospheric Aerosol and Gas Experiment
SEDI: Symmetric Extremal Dependence Index
SLCF: Short-lived climate forcers
SUBV:Solar Backscatter Ultraviolet Radiometer
TOMS: Total Ozone Mapping Spectrometer
WHO:World Health Organization
WMO: World Meteorological Organization

1.2 IAGOS - In-Service Aircraft for a Global Observing System

Contact: Andreas Petzold, a.petzold@fz-juelich.de

1.2.1 Overview

The In-service Aircraft for a Global Observing System (IAGOS; http://www.iagos.org) is a distributed research infrastructure that operates at global-scale to monitor atmospheric trace gas concentrations, aerosol and cloud optical properties by using the existing provisions of the global air transportation system (Petzold et al., 2015). It complements the global observing system in addition to ground-based networks, dedicated research campaigns and observations from satellites, balloons, and ships. This monitoring infrastructure, a legacy from the former research projects MOZAIC (Measurement of Ozone and Water Vapour on Airbus In-service Aircraft; Marenco et al. (1998)) and CARIBIC (Civil Aircraft for the Regular Investigation of the Atmosphere Based on an Instrument Container; Brenninkmeijer et al. (1999)), was formally established in January 2014 as an International no-profit Association under Belgian Law (AISBL) based in Brussels. Since 2016, IAGOS is listed as an ESFRI Landmark; see https://www.esfri.eu/ for the latest Roadmap Report.

One of the main technical improvements over the previous MOZAIC project is the transmission of data right after the aircraft lands. IAGOS main objective, besides providing a global-scale coverage of the essential climate variables (Bojinski et al., 2014) on a day-to-day basis and a set of complex observations with a reduced coverage, is also to deliver a validated (though not calibrated) data set within 3 days, on a “best effort basis”. The near real-time processing and data transmission capabilities of IAGOS will permit the measured data stream to also be included in operational processing, similar to that implemented for the numerical weather prediction models, which assimilate the aircraft measured meteorological data in near real-time. Currently, O_3, CO and H_2O measurements are available in near real time and the routine use of the IAGOS data has been particularly useful to independently validate the Copernicus Atmosphere Monitoring Service (CAMS; visit https://www.copernicus.eu/en/services/atmosphere for details) and the carbon cycle (for CO_2 emission verification) models. A cooperation with the aviation industry and instrument developers is in place to develop strategies to obtain better observation of ice particles and dust, including volcanic ash, together with their interactions.

The IAGOS infrastructure is set up as two complementary pillars:

- The IAGOS-CORE component currently deploys autonomous instruments on six long-range aircraft operated by international airlines, which make continuous, global-scale and daily measurements of temperature, water vapour, reactive gases (O_3. CO, NO_x), greenhouse gases (e.g. CO_2, CH_4), aerosol and cloud particles.

- The IAGOS-CARIBIC component, which deploys one heavily modified cargo container currently equipped with 19 instruments to measure numerous trace gases, aerosol and cloud parameters. The measurements occur once per month during four inter-continental flights.

At present (2019) 7 aircraft are equipped with IAGOS-CORE instrumentation and one aircraft carries the IAGOS-CARIBIC container. At the end of its construction phase, IAGOS aims to be operational on a fleet of 15 passenger aircraft.

The IAGOS central database hosts more than 60000 flight observations since August 1994. The database includes data from the former projects MOZAIC (38494 flights from August 1994 to December 2014) and CARIBIC, and IAGOS data since July 2011. The IAGOS data access is handled by an unrestricted access policy via the IAGOS Data Portal at www.iagos.org. The full compliance of the IAGOS Data and Services management with FAIR principles (Wilkinson et al., 2016) will be developed in the course of the upcoming EU H2020 ENVRI-FAIR project.

An exhaustive list of publications using IAGOS observations is available at http://www.iagos.org/scientific-publications/. A short list of major publications is provided below showing methodologies which have been developed for MOZAIC / IAGOS data for satellite validation together with some recent examples:

- Representativeness of the IAGOS airborne measurements in the lower troposphere (Petetin et al., 2018) and upper troposphere and lowermost stratosphere (Eckstein et al., 2017)
- Extending methane profiles from aircraft into the stratosphere for satellite total column validation (Verma et al., 2017)
- An analysis of high ozone events over India as seen by MOZAIC and IASI (Tocquer et al., 2015)
- Consistency of tropospheric ozone observations made by different platforms and techniques in the global databases (Tanimoto et al., 2015)
- Validation of nine years of MOPITT V5 NIR CO data (de Laat et al., 2014)
- Climatology of pure tropospheric profiles and column contents of ozone and carbon monoxide (Zbinden et al., 2013)

- A global climatology of upper-tropospheric ice supersaturation occurrence inferred from the Atmospheric Infrared Sounder calibrated by MOZAIC (Lamquin et al., 2012)
- Validation of six years of SCIAMACHY carbon monoxide observations using MOZAIC CO profile measurements (de Laat et al., 2012)
- Observations of the 2008 Kasatochi volcanic SO2 plume by CARIBIC aircraft DOAS and the GOME-2 satellite (Heue et al., 2010; Heue et al., 2011)

1.2.2 Satellite column validation

De Laat et al. (2012; 2014) developed a methodology to obtain profile observations during aircraft descent and ascent. For comparing measurements from satellite instruments, IAGOS vertical profile data are converted into partial columns by the procedure illustrated in Figure 1.2.1 Since IAGOS take-off and particularly landing profiles extend over several hundreds of kilometers, the satellite measurements for a given spatial area are averaged within a certain time interval. IAGOS measurements falling within this spatial-temporal "area" are considered collocated observations and are simply averaged as well.

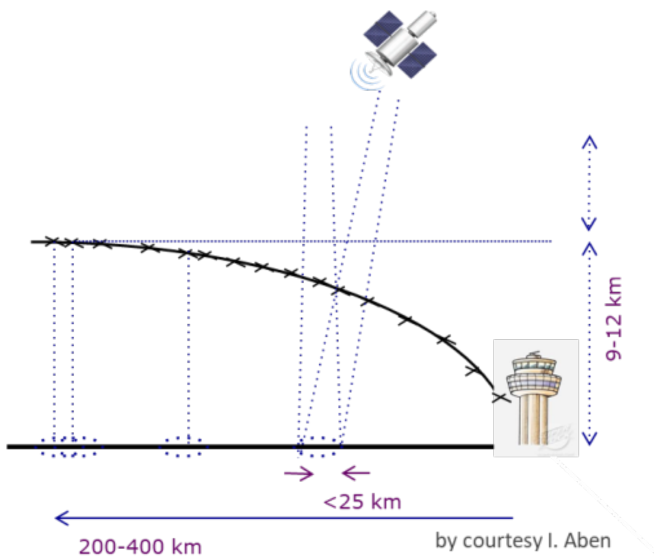


Figure 1.2.1: Schematic for converting MOZAIC/IAGOS profiles into column values.

In order to ensure that the IAGOS profile measurements are representative for a significant part of the atmosphere, those profiles are selected that start below 800 hPa up to 300 hPa at least. The missing partial column above the highest altitude beyond IAGOS data, is estimated from model results (de Laat et al., 2012), from the a-priori satellite instrument information (de Laat et al., 2014), or from climatological values (Zbinden et al., 2006; Zbinden et al., 2013; Verma et al., 2017). As an example for CO profiles, de Laat et al. (2012, their Fig. 1) showed that the contribution of the above-MOZAIC/IAGOS profile sub-column contributes less than 15–20% to the total column. Figure 1.2.2 and Figure 1.2.3 show examples of CO column concentration validations by MOZAIC/IAGOS, demonstrating its capability for satellite validation applications.

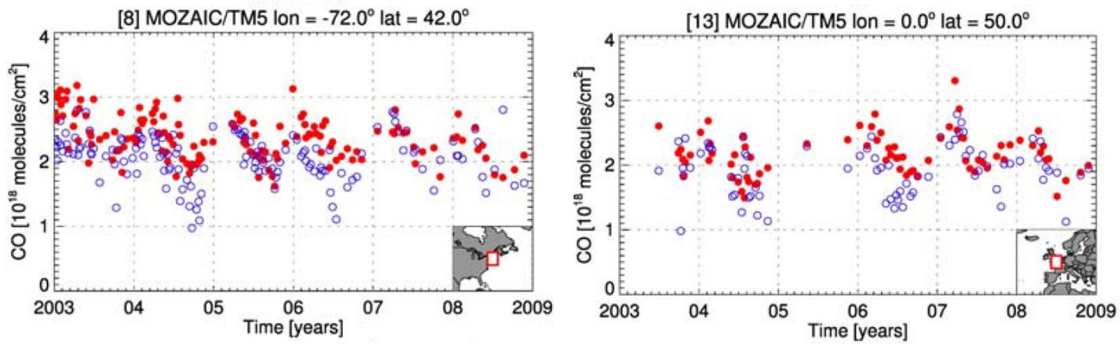


Figure 1.2.2 *Comparison of SCIAMACHY (open circles) and MOZAIC (filled red circles) CO data for the period 2003 to 2008; the region is marked in the inserted graphs. A full set of validation plots is given in Fig. 3 of de Laat et al. (2012).*

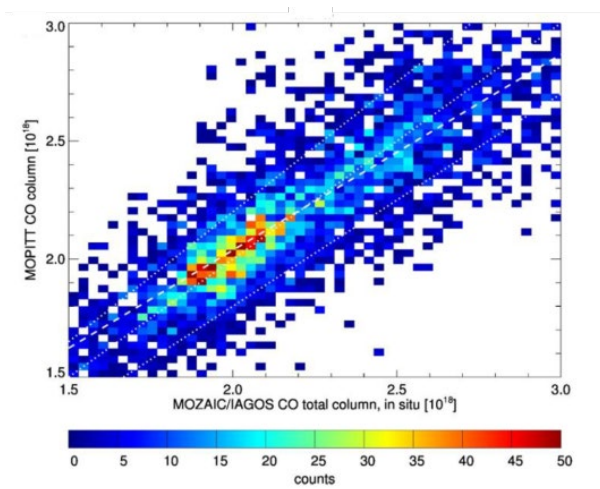


Figure 1.2.3: *Comparison of all MOPPIT and MOZAIC/IAGOS CO column total columns for the period 2002 to 2010 (de Laat et al., 2014).*

1.2.3 Satellite instrument calibration

MOZAIC water vapor data measured in the upper troposphere and in the lowermost stratosphere at cruise altitude, were used to calibrate the Atmospheric Infrared Sounder (Lamquin et al., 2012). In this study, the authors used data from MOZAIC together with the Atmospheric InfraRed Sounder (AIRS) relative humidity measurements and cloud properties to develop a calibration method to estimate ice supersaturation occurrence frequency. This method first determines the occurrence probability of ice super-saturation, detected by MOZAIC, as a function of the relative humidity determined by AIRS.

1.2.4 Representativeness of IAGOS data

In addition to the conversion of vertical profile data into column data from collocated observations, the representativeness of MOZAIC/IAGOS-CORE observations in the free troposphere were investigated by Petetin et al., (2018). In this study, the authors compared vertical profile observations of CO and O_3 in the lower troposphere with nearby surface stations from the local Air Quality monitoring network and more distant regional surface stations from the Global Atmospheric Watch (GAW) network. Based on 11 years of data from 2002 to 2012 over Frankfurt, Germany, Petetin et al. (2018) demonstrates that MOZAIC/IAGOS-CORE observations in the lowest troposphere can be used as a complement to surface stations to study the

air quality in/around the agglomeration, providing important information on the vertical distribution of pollution.
A similar study, focusing on the representativeness of aircraft data in the upper troposphere and lowermost stratosphere, was conducted comparing IAGOS-CARIBIC observations and EMAC model simulations (Eckstein et al., 2017). In this study the authors investigated to which extent such climatologies are representative of the true state of the atmosphere. Climatologies were considered relative to the tropopause in mid-latitudes (35N to 75N) for trace gases with different atmospheric lifetimes. Using the chemistry–climate model EMAC, modelled trace gases were sampled along CARIBIC flight tracks and representativeness was then assessed by comparing the CARIBIC sampled model data to the full climatological model state.
Both studies revealed that this data can be used to provide ground-truth information for satellite products also for long-term validation on a climatological basis.

1.2.5 Future plans for satellite validation activities

Since 2015, IAGOS-CORE provides vertical profiles of NO, NO_x and NO_2 over the visited airports. An example of available profiles of nitrogen oxides over Frankfurt airport is shown by Berkes et al. (2018). As a first application, comparisons of NO_2 profiles from IAGOS and OMI are currently investigated in a joint effort of Forschungszentrum Jülich and KNMI. The use of IAGOS NO_2 for the new TROPOMI instrument flying on the Sentinel 5 Precursor mission will be prepared, based on the results of the OMI-IAGOS study.

Jointly with ENVRI^{plus} Task 1.2 and Task 2.3, IAGOS project developed a new technology to routinely measure in-situ aerosol light extinction evaluating its potential application for future satellite validation efforts (Bundke et al., 2016). Following the ENVRI^{plus} Task 1.2, the synergy between RIs ACTRIS and IAGOS is strengthened making efforts joint research activities towards collocated measurements of ACTRIS Lidar and IAGOS light extinction measurements for the development of combined data sets for satellite validation applications are prepared. First successful studies have been conducted in August 2015 during a separately funded aircraft field study over the Baltic Sea. (Perim de Faria et al., 2017).

The ultimate IAGOS project validation strategy concept for satellites carrying active sensing payload like Lidar is depicted in Figure 1.2.4. While the spacecraft is carrying the active instrument on its orbit and provides global maps of target products like aerosol optical extinction or cloud coverage, the fleet of IAGOS aircraft equipped with instrumentation targeting the same property builds up global in situ data maps for the respective product. Satellite validation can then be conducted either for collocated measurements (Figure 1.2.4, left panel) or for global maps on a climatological basis (Figure 1.2.4, right panel). Once the IAGOS products for the target properties are available, respective validation activities will be prepared.

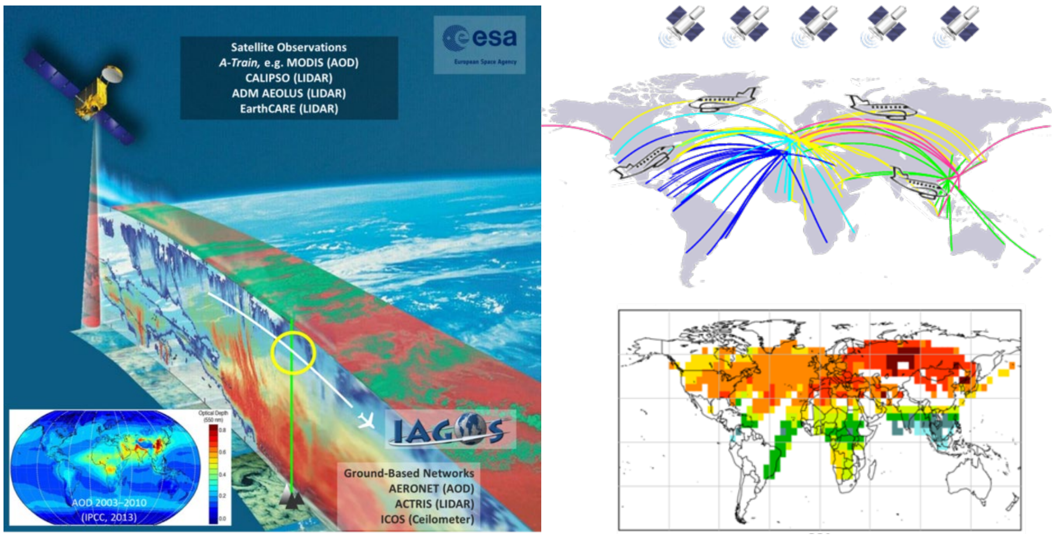


Figure 1.2.4 Sketch of the final concept

IAGOS currently available data products are listed in Table 1.2.1, 1.2.2 and 1.2.3

Table 1.2.1 AIRS Level 2 Data Products 1[&]

AIRS L2 product	IAGOS
Temperature	P1_Temperature
Relative Humidity with respect to ice; ice-supersaturation	P1_RHice
Ice cloud fraction	P1_BCP
Ozone profiles	P1_O3
Carbon monoxide profiles	P1_CO
Methane profiles	P2b_CH4

[&] access at https://airs.jpl.nasa.gov/data/products

Table 1.2.2 Sentines5P/TROPOMI Level 2 Data Products[§]

TROPOMI Product	Main Parameter	Developers	Status	IAGOS
UV Aerosol Index	aerosol index	KNMI	Released	P2e_Ext
Aerosol Layer	mid-level pressure	KNMI	2019	P2e_Ext
Carbon monoxide	total column	SRON	Released	P1_CO
Cloud	fraction, albedo, top	DLR	Released	P1_BCP
Formaldehyde	total column	BIRA-	Released	N.A.
Methane (CH_4)	total column	SRON	Dec 2018	P2d_CH4
Nitrogen oxide	total column	KNMI	Released	P2b_NO2
Ozone profiles	total and tropospheric	KNMI	2019	P1_O3
Sulphur dioxide	total column	BIRA-	Released	N.A.
Ozone (O_3)	total column	DLR	Released	N.A:

Tropos. ozone (O_3)	tropospheric column	DLR	2018	P1_O3

[§] access at http://www.tropomi.eu/data-products/level-2-products

Table 1.2.3 EarthCARE Level 2 data products[$]

EarthCARE Product	IAGOS parameter
Vertical profiles of cloud ice and liquid water content and effective particle and droplet size, cloud top and base heights, including multilayer clouds, fractional cloud cover and overlap	P1_BCP
Vertical profiles of aerosol extinction and boundary layer heights and aerosol type	P2e_Ext
Synergistically retrieved 3D cloud and aerosol scenes	P1_BCP + P2e_Ext
Observed top-of-the-atmosphere broadband radiation and radiative properties (top of the atmosphere and surface fluxes and vertical heating profiles) estimated from the retrieved cloud and aerosol fields	N.A.

[$]access:https://www.esa.int/Our_Activities/Observing_the_Earth/The_Living_Planet_Programme/Earth_Explorers/EarthCARE/Data_overview

1.3 ICOS Integrated Carbon Observation system (ICOS)

Contact:

1.3.1 Overview

The Integrated Carbon Observation System (ICOS) is a European-wide research infrastructure, measuring and monitoring the atmospheric budget of greenhouse gases (GHG) in Europe and neighboring regions. ICOS consists of harmonized network of long term observation sites for atmosphere, biosphere and ocean domains. Currently, the ICOS station network includes more than 130 stations in 12 countries. This large network is centrally coordinated through the Head Office based in Finland, the central data portal (ICOS Carbon Portal) and central facilities including atmosphere, ecosystem and ocean thematic centers and the central analytical laboratories. A map of the ICOS station network can be observed in Figure 1.3.1.

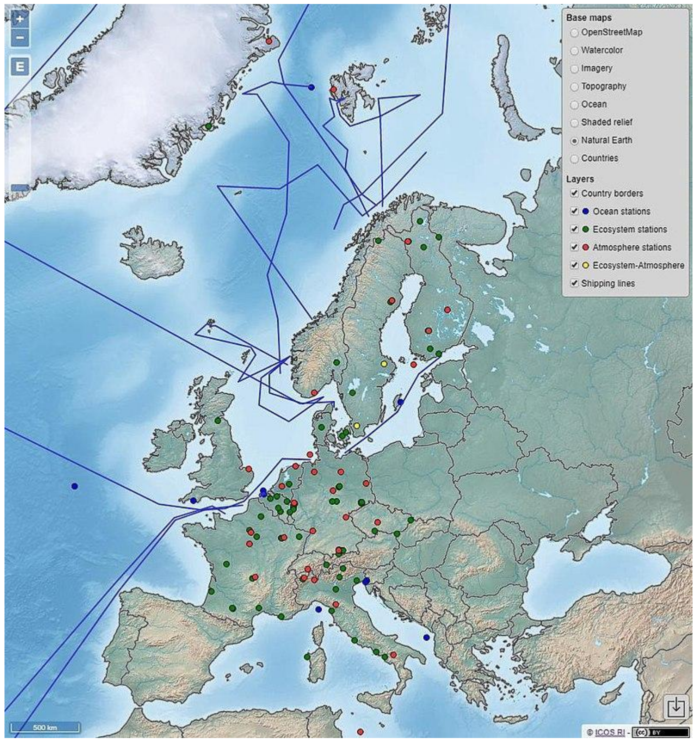


Figure 1.3.1 Network of ICOS stations. [1]

ICOS provides important long term observations, necessary to understand the current state and predict the future behavior of the global carbon cycle and GHG emissions. It aims at monitoring the effectiveness of carbon sequestration and GHG emission reduction activities on global atmospheric composition levels, including attribution of sources and sinks by region and sector.

The added value impact of ICOS infrastructure allows an enhanced visibility and dissemination of European GHG data and products that are long-term and carefully calibrated. ICOS serves as a backbone for all users, engaged in development of data assimilation models of GHG sources and sinks (inverse modeling). It delivers near-real time observations quantifying the uncertainty associated with the results, i.e. due to the use of several different models and/or using different methodologies.

ICOS is a state-of-the-art facility for the European research community and significantly contributes to the European share of global GHG observations under such international programs as GEO, WMO-GAW and GTOS. ICOS program makes Europe a top global key player for GHG in *in-situ* observations, data processing and user-friendly access to data products for validation of remote sensing products, scientific assessments, modelling and data assimilation.

ICOS products actively contribute to satellite remote sensing data assimilation in the frame of Copernicus Atmosphere Monitoring Service (CAMS) program through two contracts, CAMS84 and CAMS26 (Fig. 2).

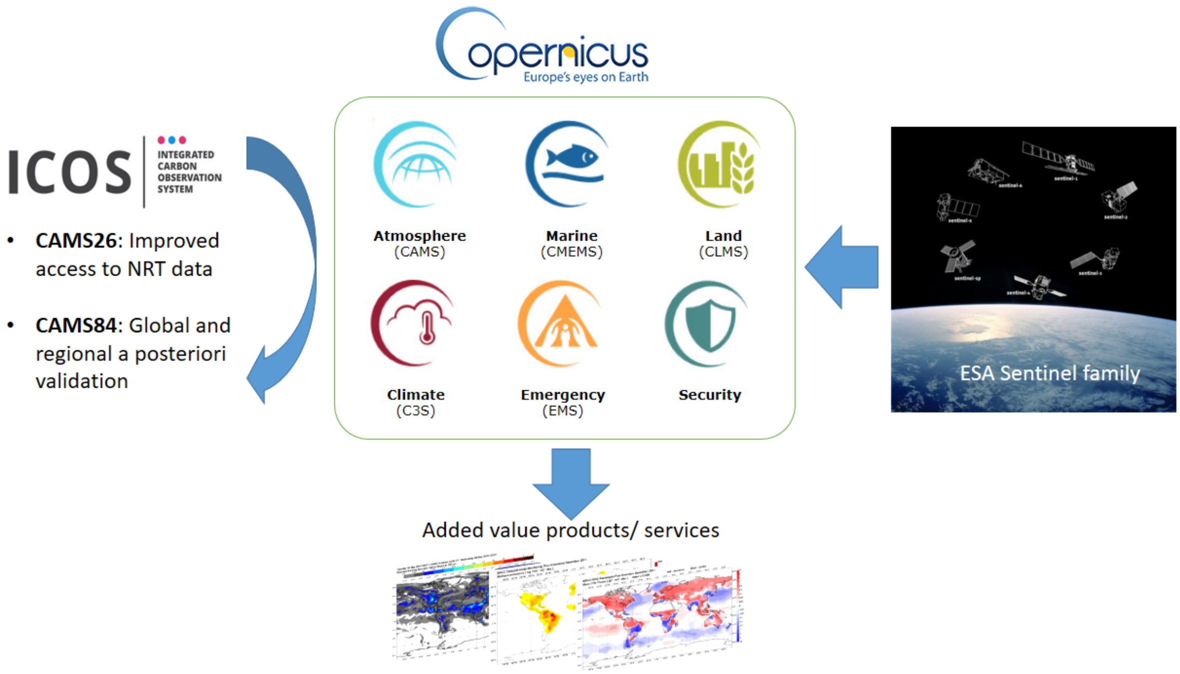


Figure 1.2.2 Structure of Copernicus Information services with the contribution of ICOS to it

1.3.2 Participation of ICOS in the CAMS84 project

At the Earth's surface, aerosols, ozone and other reactive gases such as nitrogen dioxide determine the quality of the air around us, affecting human health and life expectancy together with ecosystems health and the fabric of the built environment. Stratospheric ozone distribution influences the amount of ultraviolet radiation reaching the surface. Dust, sand, smoke and volcanic aerosols affect visibility and then pose a threat to transportation security, solar power generation, clouds and rainfall lifetime and formation and remote sensing imaging by satellite of land, ocean and atmosphere. To address these environmental issues, there is a need for data and processed information. CAMS has been developed to meet these needs, aiming at supporting policymakers, business and citizens with enhanced atmospheric environmental information.

The following operational services will be delivered:
a) Daily production of real-time analyses and forecasts of global atmospheric composition
b) Reanalyzes providing consistent multi-annual global datasets of atmospheric composition with a stable model/assimilation system
c) Daily production of real-time European air quality analyses and forecasts with a multi-model ensemble system
d) Reanalyzes providing consistent annual datasets of European air quality with a frozen model/assimilation system, supporting in particular policy applications
e) Products to support policy users, adding value to "raw" data products in order to deliver information products in a form adapted to policy applications and policy-relevant work
f) Solar and UV radiation products supporting the planning, monitoring, and efficiency improvements of solar energy production and providing quantitative information on UV irradiance for downstream applications related to health and ecosystems

g) Greenhouse gas surface flux inversions for CO_2, CH_4 and N_2O, allowing the monitoring of the evolution in time of these fluxes
h) Climate forcing from aerosols and long-lived (CO_2, CH_4) and shorter-lived (stratospheric and tropospheric ozone) agents

CAMS can be used to provide four types of products/services:

1. Real-Time Global Products: The operational real-time analyses and forecasts from the global CAMS data assimilation and forecasting system, which is run by the Global Service Provider. These analyses and forecasts can be produced at least daily and include 3-dimensional fields of aerosols, chemical species, and greenhouse gases with a temporal resolution of at least 6 hours.
2. Forecast-only Global Products: the outputs of a global CAMS forecasting system that is based on the system used to produce the Real-Time Global Products but without the assimilation of observations of atmospheric composition. The forecasts are produced at least daily and include 3-dimensional fields of aerosols, chemical species, and greenhouse gases with a temporal resolution of at least 6 hours.
3. Global Reanalysis Products: the outputs of a reanalysis from the global CAMS data assimilation and forecasting system, which is being run by the Global Service Provider. The reanalysis will cover the period between 2003 onwards and provide analyses and forecasts every 12 hours of 3-dimensional fields of aerosols, chemical species, and greenhouse gases with a temporal resolution of at least 6 hours.
4. Regional Products: the outputs of analyses and forecasts from the regional CAMS data assimilation and forecasting systems, which are run by the Regional Service Provider. The Regional Products consist in the first place of real-time analyses and forecasts. The regional CAMS data assimilation and forecasting systems will comprise at least seven individual systems as well as their model ensemble products. These analyses and forecasts will be produced every 24 hours and include 3-dimensional fields of aerosols and chemical species with a temporal resolution of 1 hour. The Regional Products also include the outputs from interim re-analyses based on fast-track in-situ observations and reanalyzes based on fully validated in-situ observations. Outputs from these re-analyzes consist of analyses of chemical species and aerosols with a temporal resolution of 1 hour and will be provided on an annual basis by the Regional Service Provider.

The CAMS84 project aims to validate the performance of those products/services, based on the atmospheric observations developed in the RIs. ICOS is contributing to the near real time evaluation of the CO2 and CH4 high resolution forecast products, and the global reanalysis currently under development.

1.3.3 ICOS participation to CAMS26 project

Operational access to NRT atmospheric greenhouse gases observations are provided to the Copernicus Atmosphere Monitoring Service (CAMS) by the ICOS Atmospheric Thematic Centre (ATC). These observations are used for verification and validation of CAMS global and regional forecasts and near-real-time analyses of global atmospheric composition and air quality. Atmospheric measurements of GHG are also used to validate and reduce climate model uncertainties. Potentially, such observations could also be directly assimilated in the global forecasting system. Consolidation and improvement of reliable preparation, transmission and quality control of near real time atmospheric ICOS data shall be performed for CAMS and other potential users.

1.4 SIOS: The Svalbard Integrated Arctic Earth Observing System.

1.4.1 Overview

The Svalbard Integrated Arctic Earth Observing System (SIOS) is a collaborative effort to develop and maintain a regional observational system for long term measurements in and around Svalbard, addressing Earth System Science (ESS) questions related to Global Change. The observing system and research facilities offered by SIOS build on the extensive observation capacity and diverse world-class research infrastructure provided by many institutions already established in Svalbard. This includes a substantial capability for utilizing remote sensing resources to complement ground-based observations. From this solid foundation, SIOS envisions a significant contribution to the systematic development of new methods and observational design in Svalbard. This knowledge can advance other observational networks in the Arctic and elsewhere. SIOS is aiming at more efficient use and better integration of the observing system based on a distributed data management system, an open access program that includes logistical support, as well as training and education activities. Working groups, task forces and other SIOS components pursue these aims in direct and structured dialogue with scientists, user groups, policy makers and other porters of societal and scientific needs. SIOS brings observations together into a coherent and integrated observational programme that will be sustained. Thus, SIOS offers unique opportunities for research and the acquisition of fundamental knowledge about global environmental change. SIOS entered the operational phase in January 2018, after a three year long interim phase (November 2014 – January 2018) and a four year long preparatory phase (October 2010 – November 2014)). Currently the consortium consists of 25 institutions from 10 countries. A map of the SIOS research infrastructures and functioning process of SIOS is depicted in Figure 1.4.1

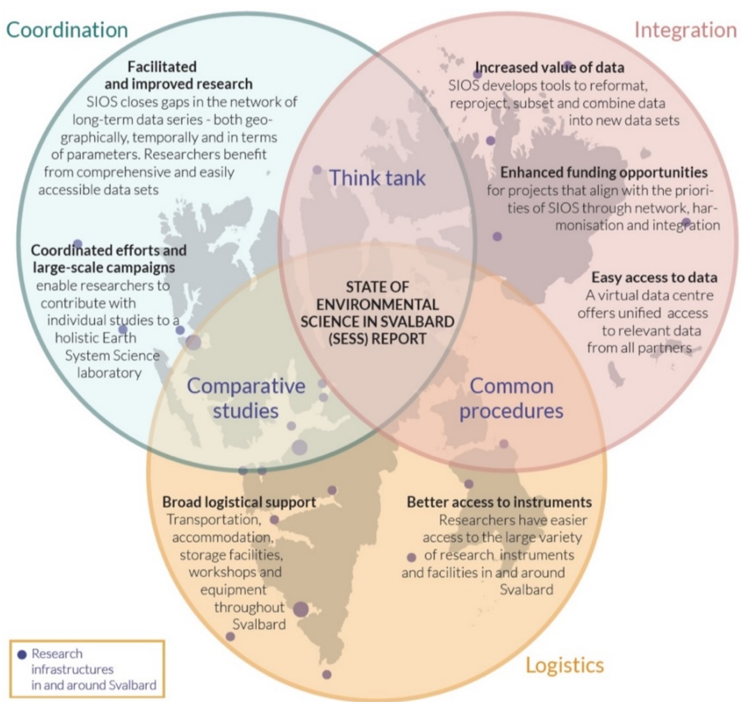


Figure 1.3.1 : Distribution of SIOS research infrastructure and functioning process of SIOS.

1.4.2 SIOS as a Copernicus Relay

The SIOS Remote Sensing Service is designed to offer to the researchers a single-point of contact for satellite information for Svalbard while drawing on the combined knowledge of the network of SIOS partner institutions.

- We coordinate commissioned data processing and make these products available via our access point.
- We advise researchers on their respective satellite data needs and also provide tailored training on remote sensing.
- As a Copernicus Relay, we aim to share our expertise with the European Commission’s Copernicus satellite programme, giving us the opportunity to showcase the application of Earth Observation data to the research community on Svalbard.

The SIOS Remote Sensing Service follows a user-driven approach where the partner institutions set the scientific needs that support the development of an extensive Observing System for Svalbard.

At the forefront of Earth Observation lies the European Commission’s Copernicus programme. The space segment of this programme consists of contributing satellite missions, known as the Sentinels, which are coordinated by the European Space Agency (ESA). All Sentinel data are freely and openly accessible online. To encourage user uptake of these data at national level, so-called Copernicus Relays have been initiated. The job of the Copernicus Relays is to promote the usage of these data through information and training activities. SIOS was successful in the bid to become one of currently three Copernicus Relays for Norway. Our aim is to encourage new user uptake by providing researchers with the necessary guidance they need to use these timely data sets.

a) SIOS training course on cryosphere remote sensing:
   A 6-day long ESA Advanced Cryosphere Training Course was conducted jointly by SIOS, NSC, UNIS and ESA in Longyearbyen, Svalbard, during 11 − 16 June 2018. This course showed the need to strengthen an active remote sensing environment in SIOS. More such events are being discussed to be conducted in 2019-2020. A similar kind of training workshop was conducted on how to use Copernicus (Sentinel-2 MSI and -3 OLCI) data in 2017. As a follow-up, the Marine Remote Sensing Workshop is being planned for the year 2019.

b) Linking Ground-based Measurements with Remote Sensing Measurements
   A major asset for Svalbard is that it is uniquely seen by, and can see, all the polar orbiting remote sensing satellites. This provides a very substantial addition to the observing capabilities of SIOS already outlined above. The combination of satellites, balloons, rockets, aircraft (manned and unmanned) and ground and marine based facilities provide data at different spatial and temporal scales, different types of detail and facilitate application of various sensing systems that in combination allow a more effective means to understand Earth System Science issues. The challenge is to integrate all these capabilities into a coordinated approach to Earth System monitoring and whilst there has been progress and significant technological advances there remains much to be done to link the different observing systems on a scale appropriate for long term observation and to make the data sets equally available.

One of the most impressive outcomes of the last International Polar Year was that all the major space agencies agreed to cooperate to make a wide range of relevant satellite data products available in common formats to the polar science community. This IPY initiative has continued and coupled with a growing number of polar relevant satellites it is transforming the research capabilities of Arctic scientists.

The users of remote sensing data for arctic science can roughly be divided into three groups:

- Users that are accustomed to utilizing earth observation data in their research. These "expert users" know the capabilities and limitations of remote sensing data as well as having the necessary software capabilities the data usage requires.
- Users that are neither well experienced in their knowledge of the capabilities of remote sensing data nor have sufficient experience in the usage of these data. For these "new users" it is essential that they receive knowledge of the capabilities and are taught how to use the necessary software. This should be done in collaboration with the "expert users" and through coursing. It is also clear that the requirement of knowledge differs significantly upon what type of remote sensing data that should be utilized.
- The wide availability of field work and ground/marine based information makes Svalbard the ideal high arctic calibration and validation site for the satellite owners.

For all types of usage of remote sensing data, it must be clear that the use must be defined by the needs of the scientists and not by the availability of data alone. There are a few areas that only can be solved by remote sensing data alone as there are only a few areas that will not have benefit of using remote sensing data.
SIOS has several roles concerning the utilization of satellite remote sensing in Svalbard science activities:

- Ensure that the member of SIOS know about the availability of the data.
- Implement training courses for new users in the different fields.
- Coordinate access to the freely available data.
- Expand the availability of satellite remote sensing data through the SIOS portal.
- Support in acquiring required commercial data.

All of these activities require coordinated efforts from the SIOS KC and of the members of SIOS that have specific knowledge and experience.

1.4.3 Satellite remote sensing data availability

For Arctic research and its input of information to Earth System Science, there are two issues that are of great importance concerning satellite remote sensing data:

- The longevity and consistency of the available data-sets. This includes long term measurements provided by EUMETSAT and NOAA as well as the broad environmentally focused missions within the Copernicus programme. These data have histories of 3 to 30 years and will persist for several decades into the future.
- In addition, specific directly scientific oriented missions with newly developed instrumentation for measurements of essential parameters in the Arctic. These missions are often "one of" missions that primarily give the data for a shorter time (4-6 years). The science missions of ESA and NASA as well as JAXA are the typical examples of this.

It is important to understand that both types of remote sensing instruments are essential for Arctic science. The long-term measurements provide information that are crucial to understand climate change. The short-term scientific measurements provide data not achievable by other means. In addition, the latter can be used to calibrate and infer additional information for the long-term missions. Now, a range of high and low altitude satellites are available or planned to provide valuable additional data, albeit not necessarily as detailed as that provided by ground-based observatories. Satellites instead provide a broader scale of observation or an integration offering valuable context.

In addition to space borne platforms, it is inherently important to make wide use of modern technology including unmanned aerial vehicles (UAVs) to understand a very high resolution short-term spatiotemporal changes over a relatively small region. Similarly, very high resolution (VHR) satellite datasets (e.g. WorldView series) will continue to provide essential geo-information over specific areas of the Svalbard e.g. glaciers, vegetation, fjords.

For all satellite measurements in the Arctic it is crucial with a broad set of Cal/Val activities. This is important for both the satellite owners and for the data users. For the satellite owners it is important for them to ensure the validity of data all over the Arctic and for users it is essential for the integration with their field measurement.

1.4.4 Specific availability of satellite remote sensing data

This information must be continuously updated. For the long-term measurements, the available parameter set is well known. For more short-term missions, like the ESA Envelope programme or the Earth observation programmes of NASA and JAXA, the data availability will be time variant. Several of these relatively short-lived programmes, like Cryosat and IceSat, provide data that are extremely relevant for Arctic research and should thus be integrated into the data series.

From experience with older satellite programmes, several of the instrumentation for specific scientific satellites will be developed and end up later as essential instruments on operational systems. This applies to several of the instruments on ESAs ENVISAT project that now are part of the Copernicus programme. Similarly, NASA and JAXA science instruments are repeated in new JAXA missions or in NOAA or USGS Landsat programmes.

For the long-term monitoring, the promised availability of the Copernicus data series is essential for understanding change. These data will be available at least until after 2030. New additional measurements for next generation Copernicus is currently being discussed. The majority of these proposals will be important for Arctic research. Functioning of SIOS to facilitate remote sensing data usage and cal/val activities amongst members and users

More than 150 EO satellites from different space agencies have acquired and continuously acquiring data over Svalbard since 2009, and will continue to provide unprecedented new data for polar research until 2030. Use of remote sensing data is well established in meteorology, cryosphere and ocean research, and many terrestrial sciences in the northern areas. Only satellites can provide systematic and consistent spatial/temporal data coverage over the polar regions, but there is significant lack of ground-based validation data, which are needed to develop appropriate retrieval algorithms and for quality control of geophysical data derived from the remote sensing data. Almost all orbits of polar-orbiting satellites pass over or at a short distance from Svalbard. This has allowed Svalbard to become the largest downlink site for polar-orbiting Earth-observing satellites. For scientific and monitoring communities these facts have a threefold positive effect: (a) the Svalbard region is excellently covered by satellite measurements, (b) research infrastructure in Svalbard offers a unique possibility for performing ground-based validation of satellite data for multidisciplinary polar research, (c) the use of satellite data has no negative impact on the environment, complementing the limited number of field stations in protected areas. In all these processes, SIOS has a significant role by promoting Svalbard as a cal/val site to conduct dedicated field campaigns and providing support via infrastructures.

SIOS Remote sensing working group (RSWG) is the nodal community which defines the user requirement document and product inventory document. User requirement document (URD) provides the comprehensive requirements of remote sensing data and products from users and members to address broad Earth system science questions. The product inventory document (PID) provides an overview of the available geospatial products for the ready use for members and users. Based on the continuous development of URD and PID, SIOS defines the strategy for providing access to prioritized remote sensing data via an access point with the help from SIOS data management system. In SIOS remote sensing data (both from satellites and air-/balloon-/rocket-borne platforms) are an important part of a comprehensive Arctic System observation and monitoring system

2 The Marine Case

2.1 Principal Marine RIs and their data products

2.1.1 Principal Research Marine Infrastructures

2.1.1.1 Euro-Argo

Euro-Argo is the European component of the global Argo programme, and is a provider and user of the associated global infrastructure. The international Argo programme (for more details, see http://www.argo.ucsd.edu/) was initiated in 1999 as a pilot project endorsed by the Climate Research Program of the World Meteorological Organization, GOOS, and the Intergovernmental Oceanographic Commission. The Argo network is a global array of more than 3500 autonomous instruments, deployed over the world ocean, reporting subsurface ocean properties to a wide range of users via satellite transmission links to data centres. Thanks to an international collaboration of more than 25 countries started in 2001, the Argo programme succeeded in setting up the first-ever global in-situ ocean observing network in the history of oceanography. In 2007, Argo reached its initial target of 3000 profiling floats. The Argo floats being battery powered with a design life of between 4 to 5 years, there is a crucial necessity to maintain the target array, by regularly deploying new floats. In that framework, 12 European countries gathered in 2008 within the Euro-Argo project with a common aim to provide an optimized and sustained European contribution to Argo by deploying 250 floats per year. After a 3-year successful preparatory phase, the Euro-Argo European Research Infrastructure Consortium (ERIC) was established in 2014 and is now able to take up this challenge by responding also to specific European interests for marginal seas, high latitudes, biogeochemical measurements and depths greater than 2000m.

The primary repository for data aggregation and dissemination is a pair of synchronized Global Data Assembly Centres (GDACs) based at Ifremer (France) and US GODAE (USA)[1]. Real-time and delayed mode data are served from the GDACs in the Argo NetCDF formats via FTP and are have an open data license. The GDACs also host aggregations of data for specific users such as a recent data collection. In

[1] http://www.argodatamgt.org/Access-to-data

addition to the GDACs Argo data are forwarded to operational users via the Global Telecommunications System (GTS), see section 5.2.1.

Argo France and the Ifremer GDAC have also introduced Application Program Interfaces (API) that enable machine-to-machine access to data, and the development of custom tools and portals on the Argo dataset. Key tools introduced are Nnidata THREDDS[2] and NOAA ERDDAP[3] tool developed by NOAA (USA). Both have a broad adoption across the USA with these services also installed on many European data holdings. Citable aggregations of data with Digital Object Identifiers assigned are are also available.

2.1.1.2 EMSO

The European Multidisciplinary Seafloor and water column Observatory (EMSO) aims to explore the oceans, to gain a better understanding of phenomena happening within and below them, and to explain the critical role that these phenomena play in the broader Earth systems. EMSO consists in a system of regional **facilities** placed at key sites around Europe, from North East to the Atlantic, through the Mediterranean, to the Black Sea. Observatories are platforms equipped with multiple sensors, placed along the water column and on the seafloor. They constantly measure different biogeochemical and physical parameters, that address natural hazards, climate change and marine ecosystems. EMSO offers data and services to a large and diverse group of users, from scientists and industries to institutions and policy makers. It is an extraordinary infrastructure to provide relevant information for defining environmental policies based on scientific data. EMSO is a consortium of partners sharing in a common strategic framework scientific facilities (data, instruments, computing and storage capacity). Formally it is a European Research Infrastructure Consortium (ERIC), legal framework created for pan-European large-scale research infrastructures.

EMSO data infrastructure has two components, the EMSODev[4] data management platform supporting real-time and delayed-mode data and forwarding of data to OceanSITES which is a primarily a delayed-mode GDAC. OceanSITES infrastructure is based on that of Argo with a GDAC hosted by Ifremer (France), an OceanSITES NetCDF exchange format, and FTP as the primary protocol for data exchange[5]. The EMSODev data management platform is designed to automated the processing and dissemination of data from EMSO modules. In addition to producing the OceanSITES NetCDF format it will potentially enable data to be shared according to OGC sensor web enablement standards in future

2.1.1.3 ICOS Ocean

ICOS Research i infrastructure has been exhaustively described in Section 1.3. The ICOS Oceanic Thematic Center currently coordinates twenty-one ocean stations from seven countries monitoring carbon uptake and fluxes in the North Atlantic, Nordic Seas, Baltic, and the Mediterranean Sea. Measuring methods include sampling from research vessels, moorings, buoys, and commercial vessels that have been equipped with state of the art carbonate system sensors. The objective is to ensure high quality measurements of greenhouse gas concentrations that are independent, transparent and reliable. In turn, this monitoring system will support governments in their efforts to mitigate climate change as well as holding them accountable for reaching their mitigation targets. ICOS-Ocean data are a delayed-mode data stream and available in the Surface Ocean

[2] https://www.unidata.ucar.edu/software/thredds/current/tds/
[3] https://coastwatch.pfeg.noaa.gov/erddap/index.html
[4] http://www.emsodev.eu/
[5] http://www.oceansites.org/data/

CO_2 Atlas[6] (SOCAT). This is available in a variety of formats and methods including; interactive data viewers, synthese and gridded files, text format, Ocean Data View[7] (ODV) format, and via ERDDAP.

2.1.1.4 EMODNet

The European Marine Observation and Data Network[8] (EMODnet) is a network of organisations supported by the EU's integrated maritime policy. These organisations work together to observe the sea, process the data according to international standards and make that information freely available as interoperable data layers and data products. EMODnet data products are available from the central portal and are split into thematic products (physics, chemistry, biology, etc). In addition to the products available from the portal EMODnet has introduced APIs to access data such as the NOAA ERDDAP tool.

2.1.1.5 SeaDataNet[9] (data aggregator)

SeaDataNet is a distributed Marine Data Infrastructure for the management of large and diverse sets of data deriving from in situ of the seas and oceans. Professional data centres, active in data collection, constitute a Pan-European network providing on-line integrated databases of standardized quality. The on-line access to in-situ data, meta-data and products is provided through a unique portal interconnecting the interoperable node platforms constituted by the SeaDataNet data centres.

Data types from the RIs can be loosely described by types:

Real-time or near-real-time data

Such data are available within 24 or 48 hours of observation time. Data assimilators such as atmosphere or ocean forecast models use these operationally. These data have a low level of automated quality control applied.

Delayed-mode data

Delayed-mode data are available after high level quality control is applied and data are documented to provide a level of quality that allow enable the data to be used in scientific applications. An examples of such applications is the calculation of upper-ocean heat content using Argo data where small biases or errors have a significant impact on results. Delayed-mode data are typically available within 12 months of the observation time.

Data products

Real-time data and delayed-mode data direct from the RIs are designed for specific purposes driven by the requirements of the RIs, these tend to be driven by their primary stakeholders. These may include specific data formats or infrastructure for the exchange of data. Data products enable broader utilization of data and may integrate data from multiple RIs into a single dataset. Examples of data products are those used for policy or governance such as biodiversity data collections. Data products can aggregate real-time, delayed-mode or both types of data depending on the requirements. Governance of the data pathways in the marine RIs is via the Global Ocean Observing System[10] (GOOS), its regional nodes such as EuroGOOS[11], and

---

[6] https://www.socat.info/index.php/data-access/
[7] https://odv.awi.de/
[8] http://www.emodnet.eu
[9] https://www.seadatanet.org/
[10] http://www.goosocean.org/
[11] http://eurogoos.eu/

JCOMM[12]. JCOMM supports JCOMMOPS[13] who coordinate and support data availability, and provide metrics for the status of each element of GOOS. Argo which includes Euro-Argo and OceanSITES which includes EMSO both have representation in JCOMMOPS.

Next section will continue by summarizing the current data availability for each RI and data product provider or integrator, and data products that integrate RI data. The section will finish by describing key recent developments that are applicable to the RIs.

2.1.2 Data availability from data infrastructures

2.1.2.1 WMO Information System (WIS)

The WIS, formerly the Global Telecommunications System (GTS), supports the forwarding of data to operational ocean data assimilations. Once forwarded to a node in the WIS that data are then forwarded to other nodes in the global network. Compared to the NetCDF formats used by GOOS systems WIF formats are much simpler with TESAC supporting only good data and restricted in size to 15kB, and BUFR allowing the inclusion of flags. Both formats have a limited number of variables (temperature, salinity, currents, with half a dozen biogeochemical variables to be introduced including dissolved oxygen and chlorophyll-A). Data are identified by the WMO number and date. The WIS is a public system with data archived by the GTSPP[14]. Since its inception the WIS has proven to be a robust and efficient way of sharing in-situ ocean data with operational agencies globally.

2.1.2.2 Copernicus - Marine Environment Monitoring Service[15] (CMEMS)

CMEMS services rely on data from in situ monitoring networks (e.g. maps, ground based weather stations, ocean buoys and air quality monitoring networks) to provide robust integrated information and to calibrate and validate the data from satellites.

The in situ networks are managed by Members States and international bodies and make data available to the services by agreement. The European Environment Agency (EEA) is leading work for Copernicus under the FP7 "GISC" project to catalogue the in situ requirements of the Copernicus services, develop frameworks and pilot agreements to ensure access to all the relevant data in a timely and sustainable way.

The development and adoption of common communication standards and adapted technology ensure the platforms interoperability. The quality, compatibility and coherence of the data issuing from so many sources, is assured by the adoption of standardized methodologies for data checking, by dedicating part of the activities to training and preparation of synthesized regional and global statistical products from the most comprehensive in-situ data sets made available by the SeaDataNet partners.
Data, value added products and dictionaries serve wide uses: e.g. research, model initialisation, industrial projects, teaching, marine environmental assessment.

---

[12] https://www.jcomm.info/
[13] https://www.jcommops.org/board
[14] https://www.nodc.noaa.gov/GTSPP/
[15] http://marine.copernicus.eu/

The current EC H2020 SeaDataNet development project SeaDataCloud[16] is moving RI data into the computing cloud, developing a virtual research environment, and new search tools to increase data discoverability.

2.1.2.3 Global Earth Observation System of Systems[17] (GEOSS)

A central part of the Group on Earth Observations (GEO) Mission is to build the Global Earth Observation System of Systems (GEOSS). GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. Networks and users can add datasets to the GEOSS data portal[18] to increase its discoverability and availability.

2.1.2.4 International Ocean Data Exchange (IODE) Ocean Data Portal[19] (ODP)

ODP aimed to implement “end to end” data management and “one-stop shop” approach. ODP can support the full range of processes including data discovery, evaluation (through visualization and metadata review) and access and delivers a standards-based infrastructure that provides integration of marine data and services across the NODC network. In fact the ODP technology is not dependent on specific thematic (discipline) content and can be scaled to other communities.

2.1.2.5 International Council for the Exploration of the Seas[20] (ICES)

Delayed mode data

The International Council for the Exploration of the Sea (ICES) is a global organization that develops science and advice to support the sustainable use of the oceans. ICES has a well-established Data Centre, which manages a number of large dataset collections related to the marine environment. ICES data and vocabularies are diverse (e.g. ship identifiers, physical oceanographic, ocean biology data) but trends to have biological delayed-mode emphasis with ICES being a key GDAC for biological data.

2.1.3 Data products

2.1.3.1 CMEMS service portfolio[21]

CMEMS was introduced in section 5.2.2. CMEMS data products are listed in the services portfolio. A the time of writing this includes 164 products including forecast results and in-situ datasets for different user requirements.

2.1.4 Active development and new requirements

So far this section has shown that there are numerous routes to access RI data in multiple formats for diverse user requirements. The is a gap in capability for data delivery linked to the diversity of the approaches and the consistency of data exposure to the web. Current efforts to improve data management are focusing on

[16] https://www.seadatanet.org/About-us/SeaDataCloud
[17] https://www.earthobservations.org/geoss.php
[18] http://www.geoportal.org/
[19] http://www.oceandataportal.org/
[20] http://www.ices.dk/
[21] http://marine.copernicus.eu/services-portfolio/access-to-products/

reducing these inconsistencies and filling any gaps in capability, making it easier of user to find and use RI data.

This section with close by reviewing recent development and active areas of development. A key upcoming conference that will set the strategy for marine data for the next decade is Oceanobs'19 as part of its data theme. At the time of writing this report Oceanobs'19 white papers are under review so this section will summarize the FAIR principles and a couple of highlights on RIs data availability from ENVRIplus theme 2.

2.1.4.1 FAIR data principles

Wilkinson et al. (2016) introduced the FAIR data principles (Findable, Accessible, Interoperable, Reusable). The FAIR principles are agnostic of discipline and the proposed adoption within the marine domain is documented in Tanhua (submitted) as part of the OceanOb'19 white papers. Adoption of the FAIR data principles is already a priority within ENVRI with the funded follow ENVRIplus project being ENVRI-FAIR that will focus on the adoption of the FAIR data principles within the RIs.

2.1.4.2 Related ENVRIplus theme 2 data availability developments

Theme 2 of ENVRIplus has developed prototypes that have the potential to increase the availability and utility of marine RI data.

- Euro-Argo have developed a data subscription service that enables users to subscribe to receive data as it becomes available

The sensor registry prototype[22] uses OGC SWE standards to harmonize sensor metadata enabling it to be shared between RIs and enhancing interoperability of metadata for users.

2.2 Marine RIs add value to satellite products

2.2.1 Chlorophyll-A case study

In situ fluorescence profiles represent a dataset of great value for probing the biology and ecology of the ocean interior. In situ fluorescence profiles and remotely sensed satellite ocean color are both proxies of the total chlorophyll-a (Chl-a) concentration. Exploiting this "sensor continuum", the merger of in situ fluorescence profiles with surface ocean color data through efficient and innovative techniques is a powerful means of enhancing the use of satellite biological data by effectively extending its reach into the ocean. Methods to do this merging are actively being developed.

Exploiting the sensor continuum between satellite and autonomous platforms is highly desirable because:

- Satellite and autonomous platforms complement one another in coverage, resolution and frequency
- The merged product can be used to improve satellite-based estimates of ocean Primary Production
- In situ fluorescence (gliders, floats, moorings, sensors on animals) are underused data
- Doing so can provide a four dimensional picture of the chlorophyll distribution in the ocean

[22] https://www.youtube.com/watch?v=4QxTZ2iiznk

Satellite ocean color data are typically calibrated using in situ surface samples analyzed by High Performance Liquid Chromotography (HPLC), both planned and opportunistic. Glider Chl-a fluorescence data is normally calibrated from samples collected specifically during glider deployment/recovery cruises, either by HPLC or fluorometry. However, the fluorescence-to-chlorophyll ratio varies depending on season, light conditions, nutrient conditions, phytoplankton physiology, and phytoplankton community composition. Therefore, purely cruise-based calibration of glider data might be insufficient to accurately estimate Chl a concentration spanning different seasons or biogeographical provinces using a single glider calibration.
Several methods have been developed for calibrating Chl-a fluorescence data against other, independent estimates of Chl-a abundance, and which may be suitable for developing the 'sensor continuum'. As an illustrative case study, here two of them are implemented on glider data collected across a full year during the UK NERC funded OSMOSIS project. This mission involved 5 gliders being deployed, often several times, from September 2012 to September 2013 at the Porcupine Abyssal Plain Sustained Observatory site (49N 16.5W) – see Table 1. As a benchmark, the glider fluorescence was also calibrated using in situ samples taken during cruises when gliders were deployed or recovered. Water samples were collected from the CTD frame Niskin bottles. Chl a concentrations were determined by filtering 250 ml of seawater onto 25 mm GF/F filters (nominal pore size 0.7 µm) and extracting pigments in 90% acetone at 4°C over a subsequent 18-20 hours period. The fluorescence of each sample was measured using a Turner Trilogy fluorometer following the method of Welschmeyer (1994). This analysis method is not quite as high a standard HPLC analysis, but still a reliable and widely used method of quantifying Chl a concentration in seawater. The resulting scale factors to convert fluorescence value to Chl-a concentration are given in Table 1 for the glider deployments.

Glider ID	Manufacturer-provided Scale factor	Re-evaluated Scale factor	R^2	# match ups
D381 cruise (September 2012)				
SG566_SepJan	0.0124	**0.0052**	0.52	64
SG533_SepJan	0.0120	**0.0053**	0.70	33
JC085 cruise (April 2013)				
SG502_JanApr	0.0134	**0.0061**	0.93	19
SG579_JanApr	0.0124	**0.0067**	0.85	19
SG566_AprSep	0.0124	**0.0051**	0.90	42
SG510_AprSep	0.0121	**0.0058**	0.85	43
JC087 cruise (June 2013)				
SG566_AprSep	0.0124	**0.0074**	0.64	135
SG510_AprSep	0.0121	**0.0070**	0.62	87
SG533_AprSep	0.0120	**0.0085**	0.61	135
JC090 cruise (September 2013)				
SG566_AprSep	0.0124	**0.0190**	0.44	36

Table 2.1.1: Calibrations of gliders against in situ samples. Taken from Rumyantseva (2016)

2.2.1.1 A neural network-based method

The first method to be illustrated is described by Sauzède et al. (2015). This method (FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) uses as input only the shape of the fluorescence profile associated with its acquisition date and geo-location. A neural network then uses this information to convert the fluorescence profiles to Chl-a profiles. The neural network is previously trained and validated using a large database including 896 concomitant in situ vertical profiles of HPLC pigments and fluorescence. These profiles were collected during 22 oceanographic cruises broadly representative of the global ocean in terms of trophic and oceanographic conditions. The geographic distribution of the 896 stations used to train and validate the neural network in the published study is shown in Figure 2.1.1. For these stations, sampling for HPLC analysis was simultaneous to the acquisition of the fluorescence profile.

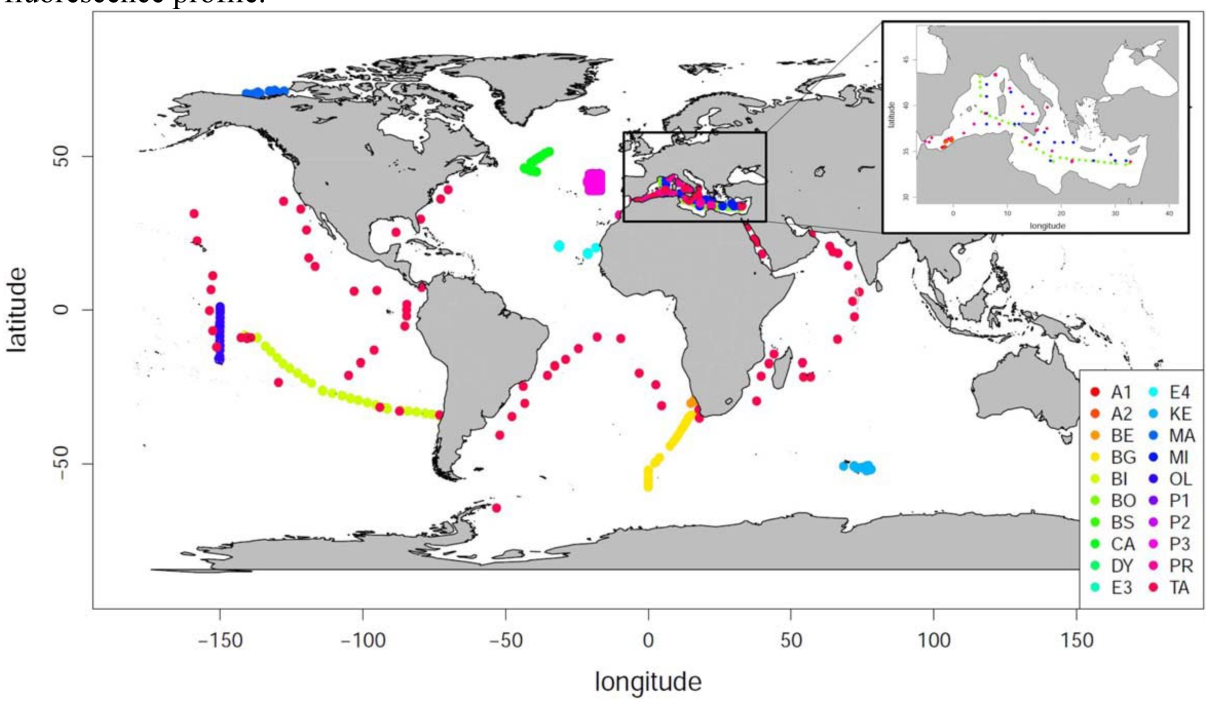


Figure 2.4.1: geographic distribution of the 896 stations used to train and validate the FLAVOR neural network approach. Reproduced from Sauzède et al. (2015).

The necessary code for carrying out the calibration and trained neural network model can be downloaded (in R or Matlab) from the supplementary materials for Sauzède et al. (2015).

The plots below illustrate how the neural network calibrated profiles compare to the water sample calibrated ones. In all cases the comparison is between concentrations averaged over the top 100m because the shape of the profile is identical for both methods; the calibrations are based on calculating a single 'scale factor' applied to the whole profile. Figure 2.1.2 shows the relative error, defined as the difference between neural network and water sample estimates, normalised by the water sample estimate. 15% of the profiles have an error of 10% or less, with 90% of the profiles having an error of 100% or less. Although scarce there are occasionally much larger discrepancies. Figure 2.1.3 shows a comparison over the annual cycle. Note that it is plotted as a single calendar year for simplicity even though data were collected from September

2012 to September 2013 (hence the jump in values in September). The seasonal variation is less pronounced in the neural network calibrated data. Finally, Figure 2.1.4 shows a scatter plot of the two calibrations. The correlation is not strong and there is considerable scatter.

In general there does not seem to be a strong match between water sample and neural network calibrated data. The PAP-SO is on the boundary between the sub-tropical and sub-polar regions of the North Atlantic. A possible explanation may be that insufficient samples from middle and high latitude North Atlantic Ocean were available to train the neural network model. Training data for the neural network are more abundant in the Mediterranean region and potentially the method may work better for the calibration of glider data in that area. It should also be remembered that while sometimes very accurate, the water sample based calibration is also sometimes much less so (see Table 1). Hence, not all of the variability between the two estimates will be due to the neural network based processing. User error should also not be ruled out. Code was downloaded from the Sauzède et al. (2015) Supplementary Material but the 'black box' of neural network' models makes it difficult to infer whether bugs have creeped into the processing when processing the fluorescence data into the format needed for analysis.

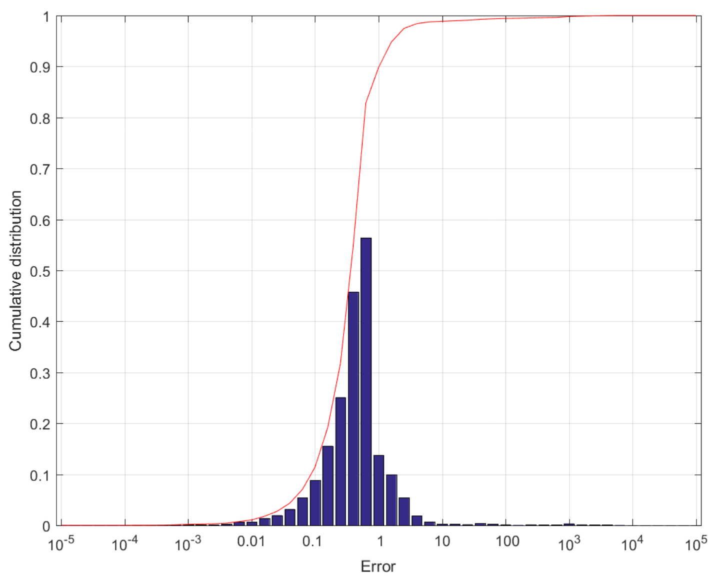


Figure 5.1.2: Error in neural network calibrated estimate (NN) relative to water sample calibrated (WS) fluorescence profiles. Error is defined as (NN-WS)/WS. Note the scale is such that an error of 0.1 is a 10% error, 1 is a 100% error and 10 is a 10-fold error. The red line shows the cumulative distribution. The histogram shows the relative distribution of errors but is to an arbitrary scale.

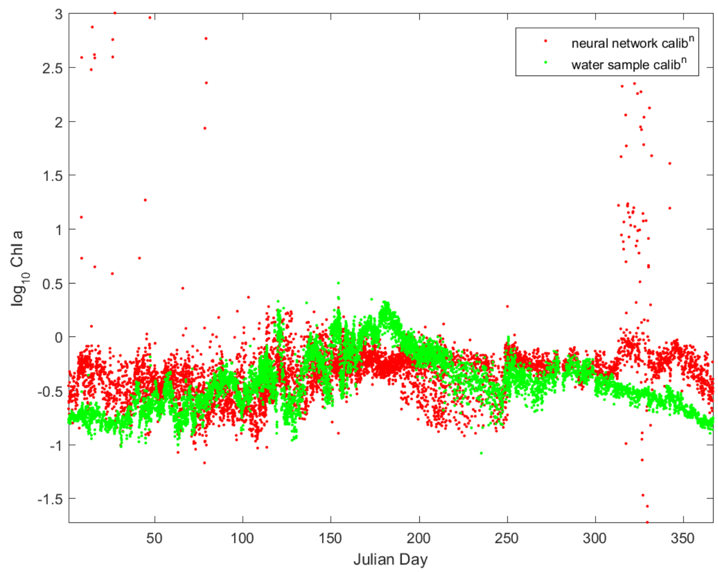


Figure 2.1.6: comparison of water sample and neural network calibrated glider fluorescence data throughout the year. Note that the data were collected from September 2012 to September 2013 but they have been plotted as a single calendar year for simplicity.

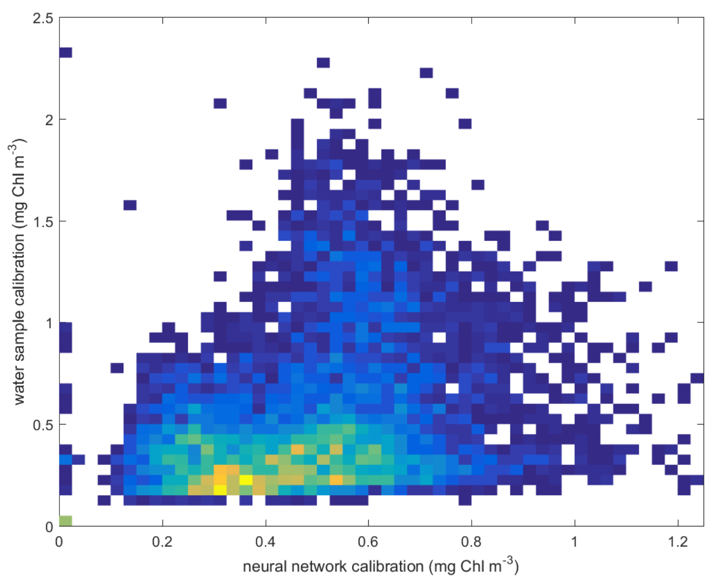


Figure 2.1.7: Scatter plot of neural network versus water sample calibrated Chl a values for the glider deployment. Some data have values outside this range but are cropped for clarity.

2.2.1.2 A satellite data merging method

The second calibration method is described by Lavigne et al. (2012). The method aims to produce "satellite-corrected" profiles by relating integrated in situ fluorescence to Chl-a stocks estimated from the satellite data. The method is based on the assumption that the near-surface Chl-a, chl_{surf} (mg m^{-3}), and the Chl *a* biomass integrated across k times the euphotic depth[23], $<chl>_{k \cdot Ze}$ (mg m^{-2}) are related (Eq. 2; Morel and Berthon, 1989; Uitz et al., 2006) such that

$$<chl>_{k \cdot Ze} = A^{*}chl_{surf}^{B}$$

where A (mg m^{-2}) and B (dimensionless as chl_{surf} is implicitly divided by 1 mg m^{-3}) are coefficients determined by regressions carried out on in situ data (Uitz et al., 2006). For the purpose of this illustration we use a value of 1 for k. To convert glider fluorescence to Chl-a concentration profiles they are multiplied by α where (Eq. 2.1.1):

$$\alpha = \frac{< chl_{k*Z_e} >}{\int_0^{k*Z_e} (FLUO(z) - dark\ counts)dz} \tag{2.1.1}$$

We show the same comparisons between water sample and satellite-method calibrated data as for the neural network illustration using values integrated over the top 100m. Figure 2.1.5 once again shows the errors relative to the water sample calibrated data. Roughly 25% of data have a 10% error or less with 98% of data having an error of 100% or less (Figure 2.1.5). The annual time-series (Figure 2.1.6) shows a better agreement in seasonal change than the neural network calibration. Figure 2.1.6 also shows the satellite data used for the calibration. Although the seasonal pattern follows that in the water sample calibrated data, there are periods (e.g. October and June) where the magnitudes differ significantly. This may reflect errors in the water sample calibration estimates but may also reflect errors in the satellite estimate of surface Chl-a. A limitation of this approach is that is impossible to calibrate the glider data when the satellite data are not available (e.g. under clouds or in winter). Figure 2.1.7 shows a direct comparison of water sample and satellite calibrated data on a point by point basis. As for the neural network approach there is significant scatter but there is some evidence of a more coherent relationship.

[23] The euphotic depth is defined as the depth at which light intensity falls to 1% of its value at the surface.

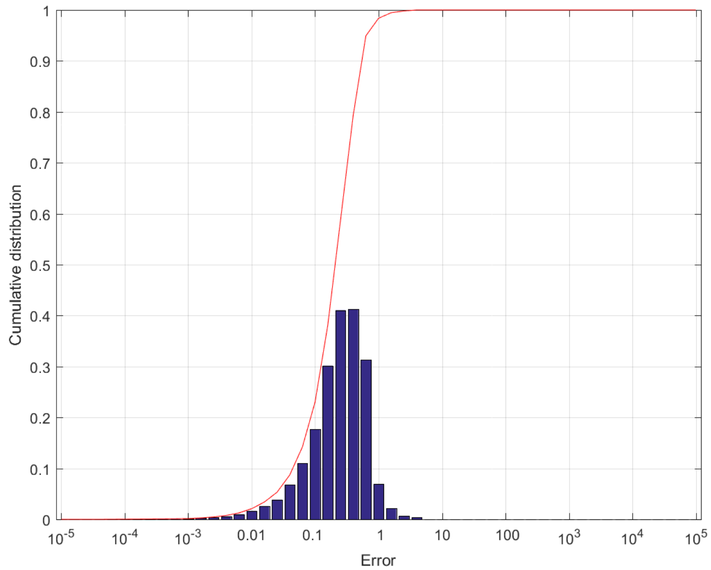


Figure 2.1.8: Error in Lavigne et al. (2012) method calibrated estimate (LS) relative to water sample calibrated (WS) fluorescence profiles. Error is defined as (LS-WS)/WS. Note the scale such that an error of 0.1 is a 10% error, 1 is a 100% error and 10 is a 10-fold error. The red line shows the cumulative distribution. The histogram shows the relative distribution of errors but is to an arbitrary scale.

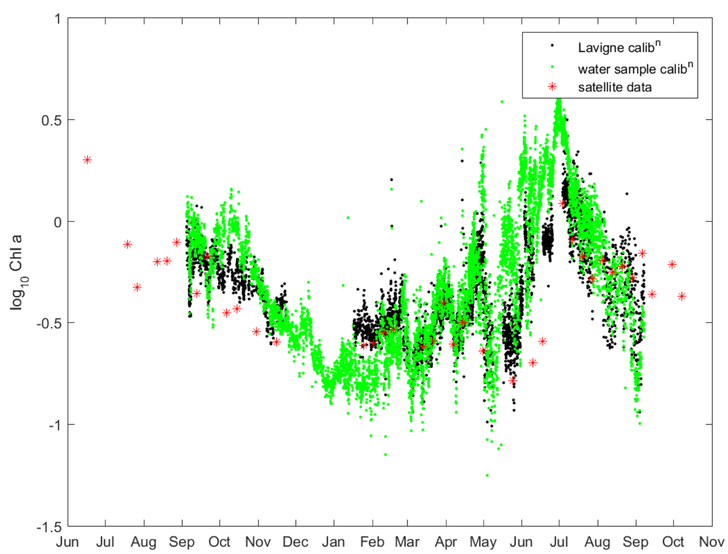


Figure 2.1.6: comparison of water sample and Lavigne method calibrated glider fluorescence data throughout the year. Note that the data were collected from September 2012 to September 2013 but they have been plotted as a single calendar year for simplicity.

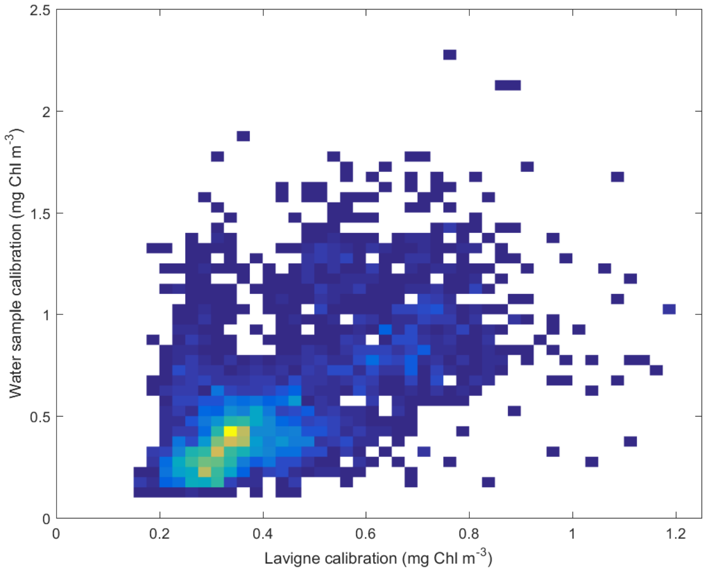


Figure 2.1.9: Scatter plot of cruise versus neural network calibrated Chl a values for the glider deployment

2.2.1.3 Use case 2: Photosynthetically Available Radiation (PAR)

PAR data from autonomous platforms like gliders can potentially be used to validate satellite daily mean PAR products. Satellites measure PAR two times per day. Using sinusoidal interpolation method, the daily cycle of PAR is reconstructed based on two measurements and daily mean PAR is estimated. On the other hand, gliders equipped with PAR sensors provide 10-12 vertical profiles per day. These data can also be used to estimate daily mean PAR.

Gliders cannot measure PAR just below the sea surface. Even if the shallowest data point was obtained 2 meters below the surface, it does not accurately represent the PAR value just below the sea surface. For instance, assuming a light attenuation coefficient (K) of 0.07 m^{-1}, then PAR at 2 meters would decrease by 13% from the value just below the sea surface ($\frac{E_{2m}}{E_0} = e^{-0.14} = 0.87$).

To overcome this limitation, and to illustrate the potential use of PAR sensors on gliders when married to satellite data, we used an exponential curve fit. A similar approach was implemented by (Thomalla et al. 2015). In this way, during daytime gliders can be used to provide several estimates of surface PAR. By way of illustration we once again use data from the OSMOSIS project.

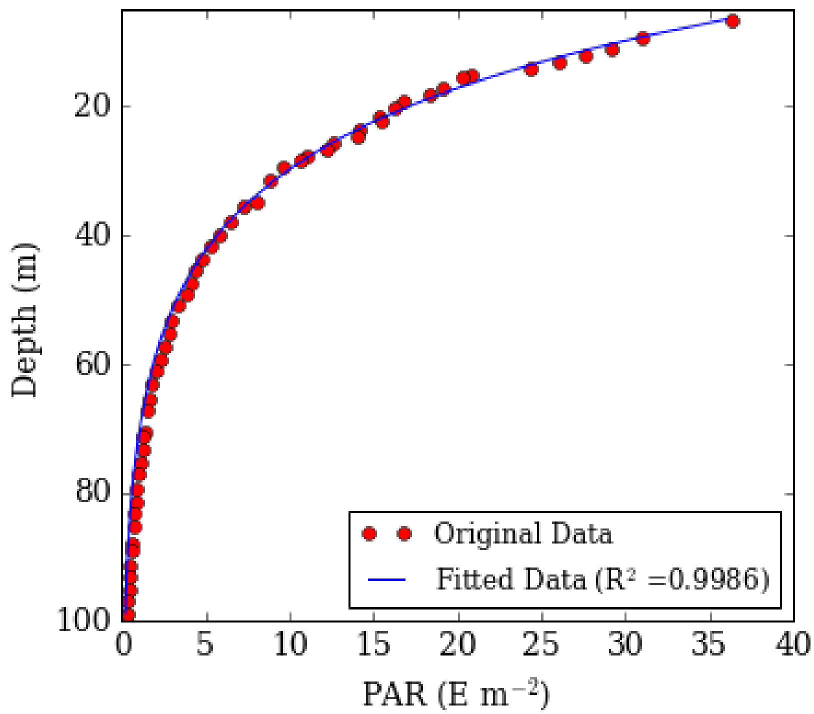


Figure 2.1.10: exponential fit to PAR data from a glider

To calculate mean daily surface PAR, we used adjusted sinusoidal interpolation (Wang et al. 2010). This approach allows reconstruction of daily cycles of PAR and, consequently, estimation of the mean value. Figure 2.1.9 shows that there is a good correlation between daily mean surface PAR obtained from satellite and from the gliders (R^2 = 0.7, slope = 0.99, intercept = 7.8).

Figure 2.1.10 shows a comparison across the full year. The daily mean satellite and glider PAR show consistent patterns with the expected strong seasonal signal. Glider estimates are generally a little below the upper envelope delineated by the satellite estimates. This is partly to the fact that gliders will provide an estimate even when the weather is cloudy. Hence the use of gliders to estimate surface PAR data may strongly complement satellite derived estimates which are prevented by cloud.

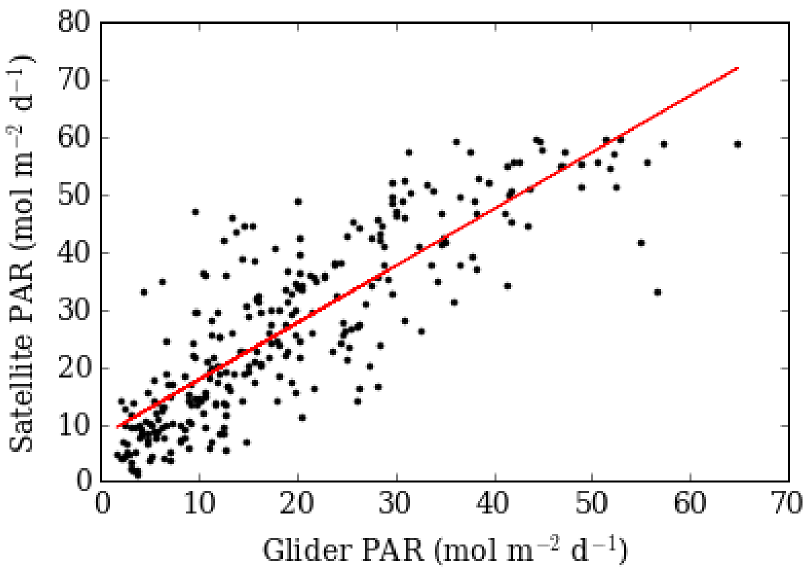


Figure2.1.11: comparison of glider and satellite estimates of daily mean surface PAR

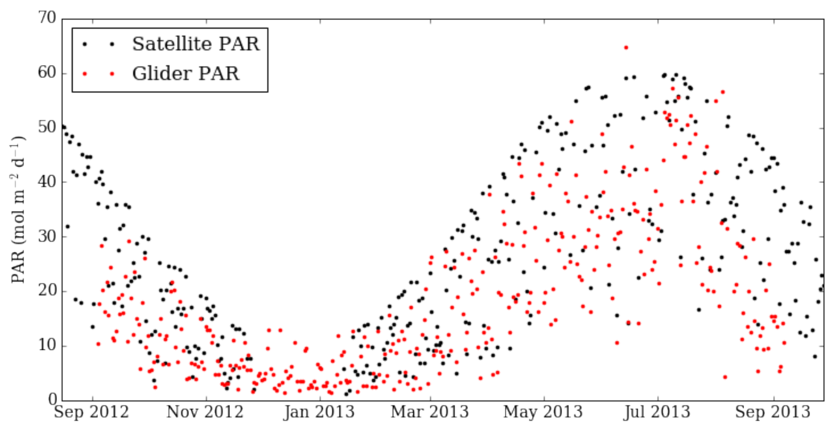


Figure 2.1.12: Comparison of satellite and glider based estimates of mean surface daily PAR through the year

2.2.2 Satellite/in situ synergies: the case of BGC-ARGO

a. Data consolidation:

The recent development of biology measurements enabled by the implementation of new sensors on board Argo floats is of high interest for space users and modeling communities as it provides a series of new datasets and methods that prove fruitful for satellite/in situ data assimilation and foster ecosystem modeling. The addressed methodologies described below are grouped by approach. Using either an analytical modus operandi, or merging the data from various sources, the presented methods put a series of applications into perspective, addressing essential issues shared by many Research Infrastructures

Several methods are developed to ensure profiles consistency. Most of the methods usually consider the deep fluorescence value as a null reference (Lavigne et al., 2012) to evaluate the offset of the instrument and then and take it into account in the output value. This provides also a technique to correct sensor derivation amongst other. Other methods (for example Mignot et al. (2011)) are based on a Chl-a concentration estimation inferred from the sole knowledge of the shape of the fluorescence profile. Profile-by-profile analysis is also useful to solve specific issues, such as the non-photochemical quenching effect (NPQ) that occurs under excessive sunlight (Maxwell and Johnson, 2000) where the measured fluorescence does not render the effective Chl-a concentration and needs to be corrected (Xing et al., 2012). For Biogeochemical-Argo floats equipped with both fluorometer and radiometer, the fluorescence data can be consolidated by using "radiometric" calibration, i.e. using concomitant radiometric measurement (that also enables Chl-a concentration estimation, see "Optical measurements" hereabove) as performed by Roesler et al. (2017) on a series of data acquired by BCG-Argo floats in various open ocean areas.

Limitations:
At a global scale, this type of analytical approach is practical and relatively reliable when there is no credible reference. Nevertheless, because a natural variability exists in the relationship between fluorescence signal intensity and Chl-a concentration, basically depending on the phytoplankton community composition, the available nutrients an d the incident light (Mignot et al., 2011), a dedicated study needs to be performed to adjust a suitable fluorescence-to-Chl-a relationship in order to get more accurate and trustful results, for each regional or local application.
Finally, this method can rapidly become fastidious, because not automatic.

b. Extend the data

Chl-a climatologies can be assembled using various measurements points and choosing a spatial and temporal resolution to extrapolate the existing data into a global pattern of the Chl-a distribution (see Conkright and Gregg (2003)). Recently a different approach, based on neural network was tested (Sauzède et al. (2015)). This method (FLAVOR for Fluorescence to Algal communities Vertical distribution in the Oceanic Realm) uses as unique input the shape of the fluorescence profile associated with its acquisition date and geo-location. The neural network is trained and validated using a large database of concomitant in situ vertical profiles of HPLC pigments and fluorescence.

Limitations:
Although both climatologies and neural network method could in theory give the possibility to expand the Chl-a picture to a larger scale than the effective sporadic measurements performed in the ocean waters, they are highly dependent on the representativeness of the samples. Indeed, applying it at a local scale, it can produce non-coherent values when the zone has specific inherent properties that no measurement point take into account (climatologies) or for which the model has not been trained (neural network).
Specifically dealing with the neural network based approach, the model is intended to use large datasets to retrieve regional or temporal trends on Chl-a climatology, but is not aimed at being used on a profile-by-profile basis (Sauzède et al., 2015). Regional scale sub-models could also be specifically created to render a particular regional behavior and produce more accurate results (e.g.: Mediterranean Sea).
For example, the application of this model on data sets collected during a cruise led by OSMOSIS (UK) consortium (Ocean Surface Mixing, Ocean Sub-mesoscale Interaction Study) highlights this limitation since many negative or extremely high Chl-a concentration values were given by the model, precisely because few samples from middle and high latitude North Atlantic Ocean were used to train it.
As a consequence, the model needs a higher number of samples to enrich its database and get better performances.

Despite a real effort made on calibration processes and analytical studies, the global fluorescence data set lacks homogeneity and standardization, which prevents from a wide efficient use of the data that it represents. It is consequently necessary to consider merged approaches (e.g. satellite) to strengthen the consistency of those databases and provide a powerful content for the scientific community to develop tools and products.

Addressing observation continuum: enhance in situ and satellite observations using their complementarity

Highlighting the complementarity
Float-derived fluorescence and satellite ocean color data are both proxies of the Chl-a concentration. Addressing observation continuum between them is highly valuable because they complement one another. Indeed, when satellites are unable to provide data, for example under cloudy conditions, floats can become the alternative (see for example Boss et al. (2008)). In addition, combining float and satellite data allows to document a 3D picture of the Chl-a distribution in the oceans by providing vertical profiles of its concentration and thus improve the satellite-based estimates of ocean primary production (Jacox et al., 2013). Finally, vertical profiles have proved to be crucial to highlight the seasonal variability in the vertical Chl-a distribution that is not observable through remote-sensing techniques (Lavigne et al., 2015).
Through efficient and innovative techniques to standardize the existing Chl-a concentration estimations, it is possible to enhance the use of both satellite and floats biogeochemical data. Such merging methods have already been developed with encouraging results in terms of Chl-a data set harmonization, real time data quality control and merged biological output.

First approach: Pairing satellite and in situ surface measurements
Boss et al. (2008) focused his study on one single float and matched the data obtained during 3 years with co-located ocean color remote sensing products. Matchup are analyzed to check the correspondence of the estimated Chl-a concentration and test the floats data reliability. This work showed that the fluorescence data proves to be consistent and that floats could thus be used as the in situ correspondents of the remote sensing devices when no satellite data is available, mainly in winter. More recently, Schaeffer et al. (2016) use and discuss another "basic" match-up analysis between in situ surface Chl-a fluorescence and ocean color satellite products from MODIS to technically validate glider measurements.

Going deeper: Integration of the whole vertical profile
With a similar approach to Boss et al. (2008) [8], Lavigne et al. (2012) uses the complementarity of the data sources to produce "satellite-corrected" profiles. The method is based on the assumption that the near-surface Chl-a and the integrated Chl-a biomass across the water column are related to each other (Morel and Berthon, 1989; Uitz et al., 2006). The method describes how the integrated Chl-a content for each vertical profile is compared to the near-surface Chl-a concentration provided by concomitant satellite ocean color measurements. Hence, thanks to the Chl-a stocks estimates based on the satellite data and using the above-mentioned relationship, the integrated fluorescence parameters over the whole water column can be adjusted and the float data gets consolidated.

Seeing wider: Multiplying the platforms
In order to understand the main drivers of the annual biogeochemical cycles in the Mediterranean Sea, Mayot et al. (2017) uses a multiplatform approach where the use of satellites, ship-based sampling and above all in situ autonomous platforms (i.e., Argo and BGC-Argo floats, bio-optical gliders) are combined. When no direct in situ comparison could be made between fluorescence and bottle measurements, the satellite-derived method described by Lavigne et al. (2012) has been applied. First tested on a large range of both fluorescence

and bottle samples, the method gave satisfying results in terms of data consistency, hence proving the complementarity and possible interoperability of remote and in situ platforms for larger data availability. Since then, this method has been largely used for glider and float calibration (Bosse et al., 2017 , Pasqueron de Fommervault et al., 2017)

Although the above-mentioned lately developed methods gave satisfying results in terms of data consistency and let envisage a potential future interoperability of autonomous platforms, still a few limitations remain that one must keep in mind, for which some enhancements are possible.

Limitations, recommendations and further extensions

NPQ (non-photochemical quenching)
Although a series of methods have been developed to retrieve the Chl-a concentration thanks to in vivo fluorescence in case of NPQ (Xing et al., 2012), a part of uncertainty remains on both remote sensing and fluorescence proxies leading to data homogeneity issues. One of the proposed way to counter this issue is to avoid coupling the data at certain hours of the day for the surface measurement, typically noon (i.e. when NPQ is not present).

Environmental conditions
Even though merging in situ and satellite estimations is a promising approach, no comparison of the data is possible when the studied zone is masked by clouds, and calibration of the data remains impossible (Boss et al., 2008). However, considering long term calibration processes ensuring data validity and consistency between fluorometers and ocean color sensors, this issue could be raised in the future.

Spatial and temporal scales
As described in Lavigne et al. (2012) , there is a compromise to adjust between the number of float to satellite match-ups and the considered temporal and spatial scales, directly impacting the merged data validity. Based on a study performed on HPLC profiles match-ups with satellites, the optimized float to satellite match-up procedure has been defined on an 8-day temporal scale with ±0.1° squares spatial resolution (as a comparison, Boss et al. (2008) uses 1-day temporal/1km spatial resolutions). This implies that the existing method do not allow a real-time merging, but a certain delay needs to be considered, which is anyway consistent with the floats profiling cycle (~10 days). Hence, balancing the accuracy of the required data, shorter time scales could be considered.

Near coastal
Due to the absence of profiling floats in the near coastal environment, no data would be available for such a merging method as proposed before. Fixed moorings could help to solve this issue by implementing a net of autonomous platforms in the near coastal area, however, considering the existing portfolio at the date, no vertical profiles would be available. In addition surface equipment is always subject to collisions with maritime engines, whereas floats get the advantage of using a stationary depth of 1000 meters, minimizing the chances of collision.

Representativeness of calibration
As mentioned before, fluorescence-to-chlorophyll ratio varies depending on season, light conditions, nutrient conditions, phytoplankton physiology, and phytoplankton community composition (Mignot et al., 2011). Therefore, due to the variability of the fluorescence properties of the phytoplankton in different

biogeographical regions or at different times of the year, global calibration processes could lack representativeness. However by re-adjusting the correction factors to be considered for different latitude ranges (Haentjens et al., 2017), the models can be tuned and permanently get more realistic.

Applications

In terms of numbers of observations, fluorescence profiles represent nowadays the most important source of in situ data for Chl-a (i.e. 61 775 profiles in the World Ocean Database 2013; Boyer et al., 2013). Considering the recent development of autonomous platforms equipped with fluorometers and the ongoing deployment of new floats, this trend is expected to increase in the near future. By augmenting the number of available data and strengthening its value, numerous applications can be put into perspective to monitor, model, and forecast future environmental trends.

- **Towards improved biogeochemical models**

An improved picture of the global chlorophyll distribution with better satellite and in situ data consistency combined with the addition of the vertical dimension is likely to enhance the performance of present biogeochemical models, taking more parameters into account and linking them for better correlation analysis. As an example, the ESTOC (Estimated State of Global Ocean for Climate Research, Japan) program, developing estimations of the global ocean physical and biogeochemical state, could benefit from such improvement in biogeochemical data availability and consistency. It is also the case for the COFS (Coastal Ocean Forecasting System, Kourafalou et al., 2015) that uses open ocean data to monitor and forecast the state of the coastal seas.

- **Climate and primary production - CMEMS/Weather, Climate & Seasonal Forecasting**

Through the enhancements allowed by the merging of satellite and in situ data, an extended set of data could be available on various points where meteorological measurements are proceeded and could then be helpful to link climate data and primary production, as well as get more accurate estimations and forecasts of phytoplankton biomass through the 3D aspect provided by vertical profiles.

- **Ecological value - CMEMS/Marine Resources**

Appraising primary production is also having an eye on the dynamics of the higher trophic levels. Such satellite and in situ data merging methods can provide key inputs for the conception of products performing an efficient monitoring and eventually forecast the phytoplankton distribution, enabling the development of enhanced ecological models (Gehlen et al. (2015))

- **Support to Marine/Maritime Spatial Planning (MSP) - CMEMS/Marine Resources**

A better understanding of the drivers inferring the ocean's primary production at adapted geographical and temporal scales thanks to better observations and enhanced models will help to create new tools and products to support marine policies and foster the management of the interactions between all the actors of the marine domain through a proper identification of the impacts of the maritime activity (European Commission - Maritime Spatial Planning [22]). As an outcome of Marine Spatial Planning, the definition of potential Marine Protected Areas is rendered more pertinent with such tools, delineating zones with permitted and non-permitted uses (IUCN (International Union for Conservation of Nature)) for sustainable environment protection and human multi-sector activity (UNESCO/IOC's Marine Spatial Planning Programme).

2.2.3 Representativity of particular sites for particular measurements

Fixed point observatories provide some of the longest time-series available to us for studying the dynamics of the ocean and are a critical tool, particularly for the study of seasonal and inter-annual trends in biology and biogeochemistry. While they clearly provide significant insights into processes in their immediate

locality, it is important to understand how typical they are of larger areas, in order to determine the applicability of results obtained at this small number of sites globally. Do they effectively represent much of the world's ocean or are we still failing to monitor the majority of the seas?
The question of how coherent ecosystems are on larger scales is a long-standing one. Although Alan Longhurst acknowledged there were similar ideas before him, the classic work is his Ecological Geography of the Sea (2006). In that book, Longhurst used properties, particularly hydrographic and chemical boundaries, to divide the ocean up into provinces. Figure 2.2.1 shows these provinces.

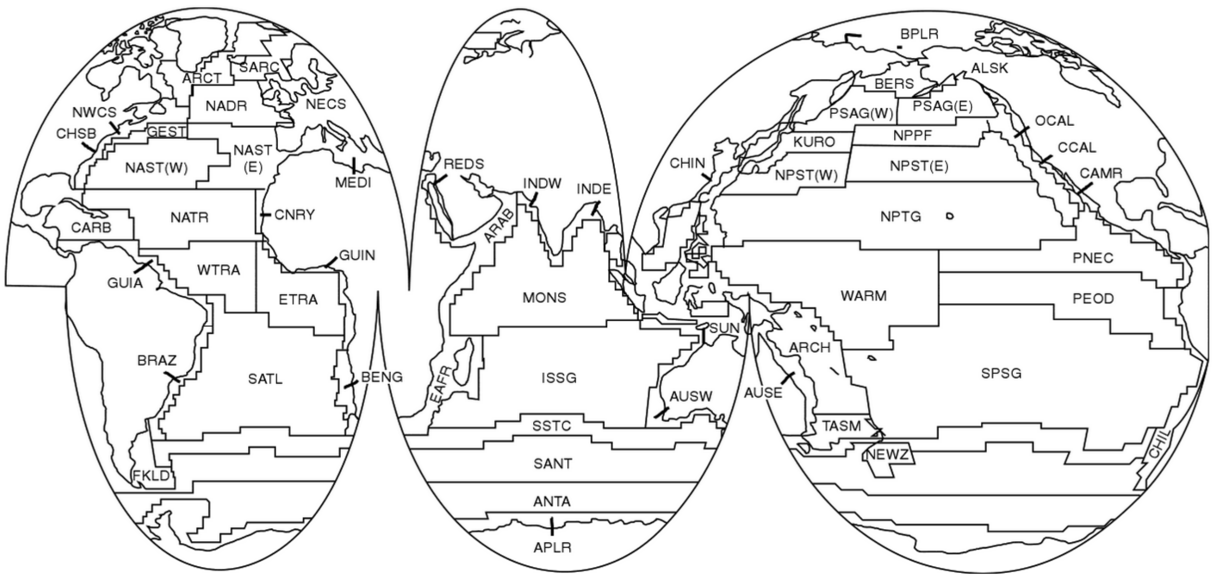


Figure 2.2.13: Ocean provinces reproduced from Ecological Geography of the Sea, Longhurst (2006)

Many papers have revisited the Longhurst provinces. Most significantly the question of the scale at which a point represents a wider area has been reconsidered from other perspectives. The Longhurst approach is to divide the ocean up into regions on the basis of properties thought to influence ecosystems, for example mixed layer depth, temperature, or oxygen. One appeal of this approach is that such data are relatively easy to obtain. However, the approach is rather reductionist, with little scope for variability related to other influences, in particular internal dynamics of the ecosystem. An alternative perspective is to address the question informed by which property is of particular interest: if we are interested in how representative a time-series station is of pH dynamics in a wider area then it has been argued that we should use temporal variability in pH to assess this. An obvious obstacle is the need for data from wider areas to allow this. After all, there would be no need to assess representativity if we already had data from many more locations. Nevertheless, models do give a means of exploring this perspective. Figure 2.2.2a shows the 'footprints' of 33 planned or existing fixed-point time series sites (http://www.oceansites.org) for temporal trends in pH. Taking the output from eight Earth System Models with the same atmospheric forcing, Henson et al. (2016) defined the 'footprint' of a fixed point station as the collection of adjoining pixels that share similar statistical properties of the time series. More specifically, other locations were judged similar to a fixed-point location if the time-series were strongly correlated and the means were less than two standard deviations (of the fixed point mean) apart. Although some fixed-point sites have considerably sized footprints for pH (for example the OOI global array at 55S 89W), the sizes are very variable. Furthermore, the shape and size of the footprints change according to the property being studied. The other panels in Figure 2.2.2 show footprints for the same time-series site locations but for other properties: SST, oxygen concentration, nitrate concentration, abundance of non-diatom phytoplankton, Chlorophyll concentration, export production and primary production. There is no consistency in the relative size of footprints at different locations for different fields. For example: Chlorophyll concentration and primary production have similar size footprints in the southeast Pacific but very different sized footprints in the subtropical North Pacific; nitrate concentration and pH are similar in the subtropical North Pacific but very different off northwest Africa.

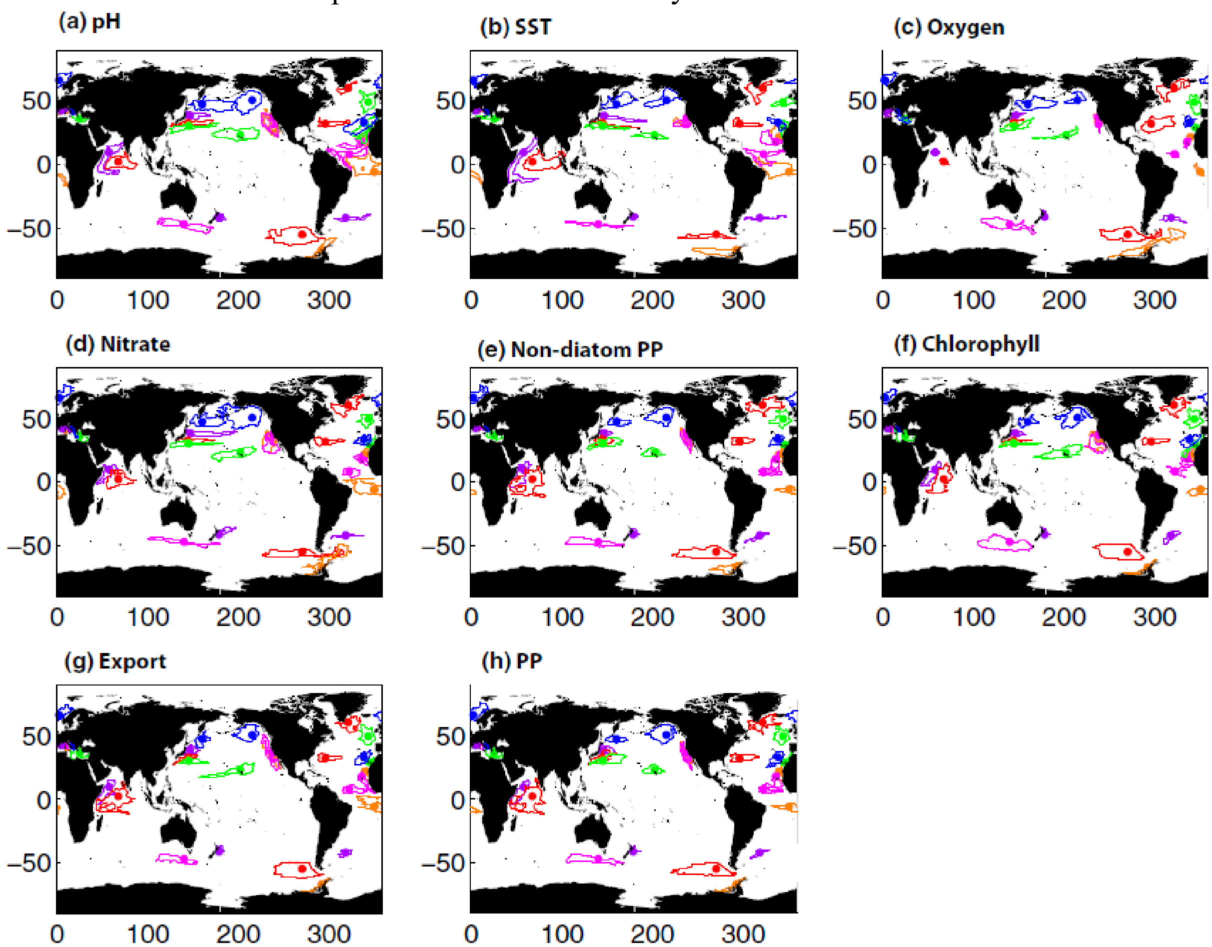
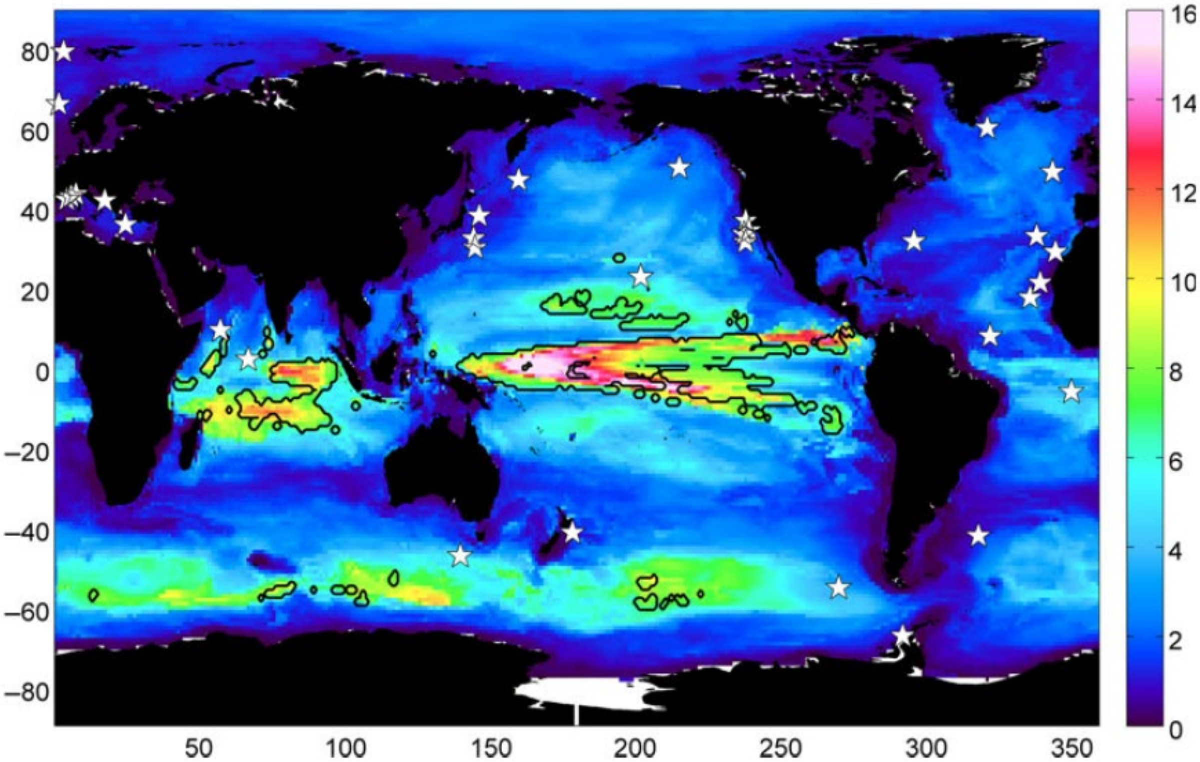


Figure 14.2.2: Each dot represents a current or planned fixed-point station. The contour in the same colour shows the extent of the fixed-point station's 'footprint'. Reproduced from Henson et al. (2016)

It is apparent in Figure 2.2.2 that the cumulative area of the footprints associated with the 33 fixed-point stations is still a small fraction of the total ocean area, regardless of the parameter under consideration. On average each footprint covers 1.1×10^6 km^2 (Henson et al., 2016). The smallest cumulative area corresponds to oxygen concentration with the total footprint area being just 9% of the total surface area (390×10^5 km^2). Even the largest (pH) is only 15% of the total area.

One possibility is that the choice of fixed-point locations is not optimal. This might not come as a surprise given that several were chosen for logistical convenience or historical reasons. Figure 2.2.3 shows the size of footprint that would correspond to making any given location a fixed-point time series site, based on the Chlorophyll concentration time series purely by way of illustration. It is noticeable that the majority of the fixed-point times-series sites are in regions with relatively small footprints. However, it is seen that the majority of points in the ocean would have similarly small footprints. Only the tropical Pacific and Indian

Oceans, and parts of the Southern Ocean, have locations with substantially larger footprints. Given the small fraction of the total ocean area covered by these regions, it would still be challenging to find 33 locations for any parameter that would cumulatively represent more than half of the global ocean.



***F**igure 2.2.3: Size of footprint in units of 10^6 km^2 for each location based on Chlorophyll concentration time-series from 8 models. Stars show locations of existing time series stations. Black contours show regions with particularly large footprints and where relatively shorter time-series are required to detect change. Reproduced from Henson et al. (2016)*

The approach used in Henson et al. (2016) is an effective means of assessing the individual and cumulative areas represented by our fixed-point stations if one has a clear idea of the level of similarity required to be in a footprint; in this case the specified criteria were the strength of the correlation required and the acceptable distance between means. An alternative is to start from the perspective of defining the number of time-series stations; if one could only afford 33 fixed-point stations, how could they best be located to represent the total ocean? Note that this is a different question to that tackled by Henson et al. (2016). There is no longer an absolute, quantifiable strength of relationship between two points for them to be considered similar. Instead, the question is, for any given point, which other points is it most similar too? This allows points to be grouped even if the relationship between them is weaker than in the Henson et al. (2016) approach.

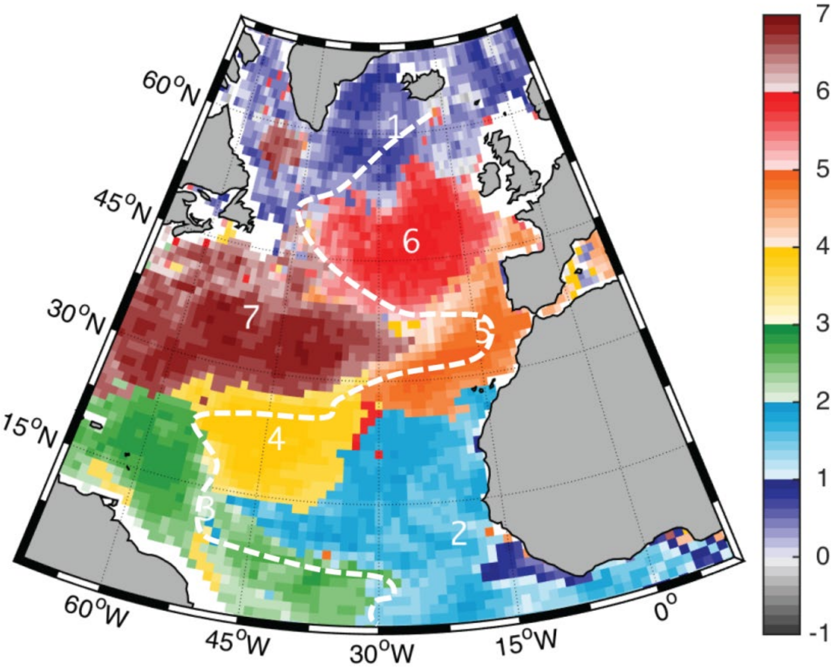


Figure 2.2.4: Clustering analysis applied to divide the North Atlantic into 7 regions on the basis of a satellite derived Chlorophyll time-series. Reproduced from Gravelle (2016)

One method to achieve this approach is clustering analysis, which can be used to divide the ocean into a number of regions decided in advance. It achieves this by making use of an iterative approach, minimising the difference between groups of points by shuffling points from group to group until a solution is found.

As an illustration of the application of clustering analysis, Figure 2.2.4 shows it applied to a time series of satellite Chlorophyll data (Monthly data spanning 1998 through 2012) from the North Atlantic, taken from Gravelle (2016). The number of regions has been chosen in advance to be seven. Each region, or 'cluster', has been ascribed a different colour, with the intensity of colour indicating the strength of correlation with the mean time-series for the whole region. It is seen that there is usually a core region of high colour intensity surrounded by a boundary of weaker colours, where the correlations are weaker. The Henson et al. (2016) approach would only have included the points of highest colour intensity in the footprint. Clustering analysis makes use of relative similarities to effectively extend the definition of footprints into areas of weaker correlations. Caution is needed because choosing an insufficient number of clusters risks grouping points even if they are not very similar in time-series characteristics. There are means of choosing the number of clusters for a given dataset (Gravelle, 2016) but a first order check can be achieved by repeating the analysis using a different number of clusters and seeing how the differences within clusters changes. The dashed line on the figure shows the North Atlantic divided into just 2 clusters. This may seem coarse but even this gives a good match to the pattern of the North Atlantic Oscillation dipole (Gravelle, 2016). Although not perfect the regions defined by cluster analysis align quite well with the dipole boundary indicating that the approach has some skill in extending the boundaries of the Henson et al. (2016) footprints.

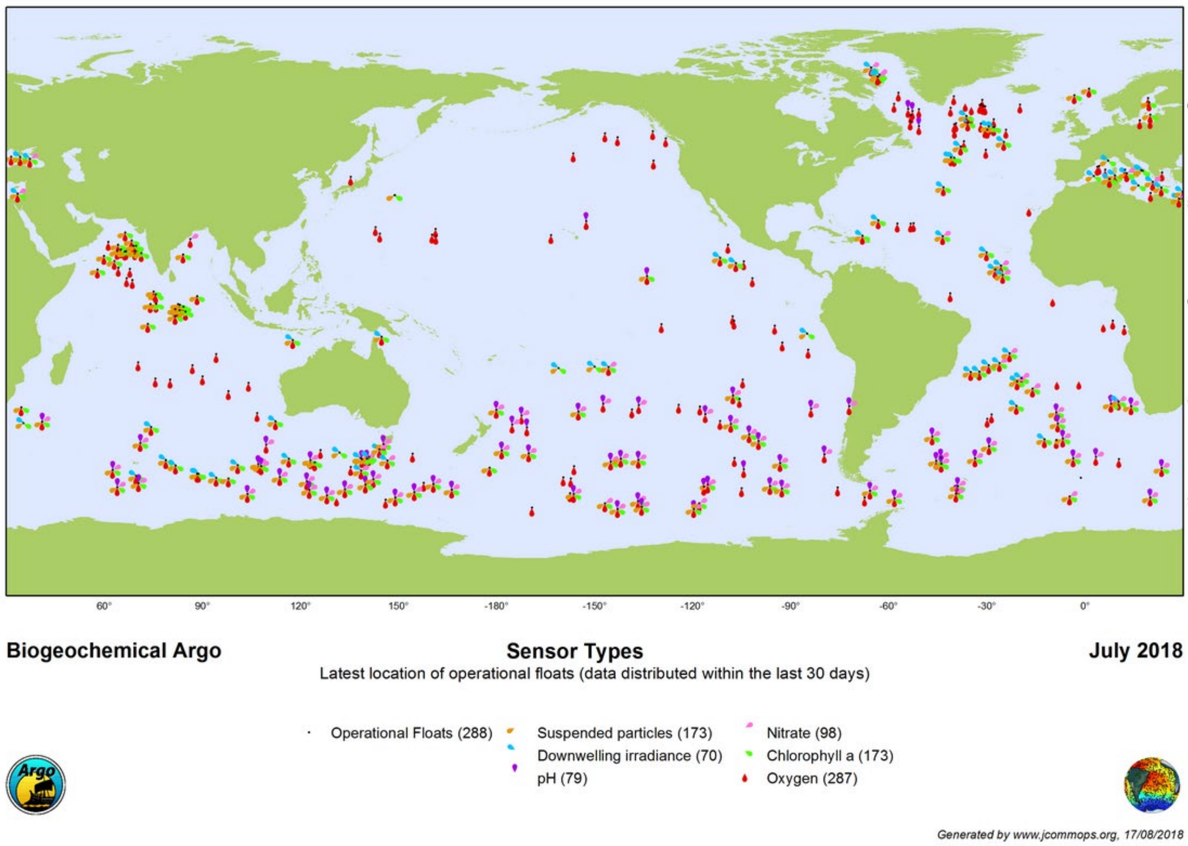


Figure 2.2.5: Locations and sensor types of Biogeochemical Argo floats as of July 2018. Reproduced from [***http://www.biogeochemical-argo.org/float-map-network-status-maps.php***](http://www.biogeochemical-argo.org/float-map-network-status-maps.php)

Looking to the future, the discussion of how representative existing and planned fixed-point stations are of the global ocean needs to be framed in the context of the growing Biogeochemical Argo programme. Figure 2.2.5 shows the current distribution of BioARGO floats. They vary according to the number and type of parameters that they detect and they remain patchily distributed. Nevertheless, they already significantly exceed the number of fixed-point stations and their numbers are growing. Fixed-point stations will remain a valuable resource for their accumulated time-series, as foci for process studies and as a platform for trialing new sensors that may currently be too power hungry for use on autonomous floats. However, the expansion of the Bio-ARGO fleet in number of floats and diversity of sensors means that we may eventually be able to move to constructing time-series for any region of the ocean by combining data from the many floats that pass through it, rather than by having to establish a fixed physical presence.

3 References

Amiridis, V., Marinou, E., Tsekeri, A., Wandinger, U., Schwarz, A., Giannakaki, E., Mamouri, R., Kokkalis, P., Binietoglou, I., Solomos, S., Herekakis, T., Kazadzis, S., Gerasopoulos, E., Proestakis, E., Kottas, M., Balis, D., Papayannis, A., Kontoes, C., Kourtidis, K., Papagiannopoulos, N., Mona, L., Pappalardo, G., Le Rille, O., and Ansmann, A.: LIVAS: a 3-D multi-wavelength aerosol/cloud database based on CALIPSO and EARLINET, Atmos. Chem. Phys., 15, 7127-7153, https://doi.org/10.5194/acp-15-7127-2015, 2015.

Antoine et al.- The "BOUSSOLE" Buoy - A New Transparent-to-Swell Taut Mooring Dedicated to Marine Optics: Design, Tests, and Performance at Sea, 2008 https://journals.ametsoc.org/doi/full/10.1175/2007JTECHO563.1

Balis, D., Koukouli, M.-E., Siomos, N., Dimopoulos, S., Mona, L., Pappalardo, G., Marenco, F., Clarisse, L., Ventress, L. J., Carboni, E., Grainger, R. G., Wang, P., Tilstra, G., van der A, R., Theys, N., and Zehner, C.: Validation of ash optical depth and layer height retrieved from passive satellite sensors using EARLINET and airborne lidar data: the case of the Eyjafjallajökull eruption, Atmos. Chem. Phys., 16, 5705-5720, https://doi.org/10.5194/acp-16-5705-2016, 2016.

Berkes, F., Houben, N., Bundke, U., Franke, H., Pätz, H. W., Rohrer, F., Wahner, A., and Petzold, A.: The IAGOS NOx instrument – design, operation and first results from deployment aboard passenger aircraft, Atmos. Meas. Tech., 11, 3737-3757, doi: 10.5194/amt-11-3737-2018, 2018.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of essential climate variables in support of climate research, applications and policy, Bull. Am. Met. Soc., 95, 1431-1443, doi: 10.1175/bams-d-13-00047.1, 2014.

Bosse et al.- Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.4319/lo.2008.53.5_part_2.2112, 2008

Bosse et al. - A submesoscale coherent vortex in the Ligurian Sea: From dynamical barriers to biological implications https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1002/2016JC012634, 2017

Boyer et al. (2013) - World Ocean Database 2013 https://data.nodc.noaa.gov/woa/WOD13/DOC/wod13_intro.pdf

Brenninkmeijer, C. A. M., Crutzen, P. J., Fischer, H., Gusten, H., Hans, W., Heinrich, G., Heintzenberg, J., Hermann, M., Immelmann, T., Kersting, D., Maiss, M., Nolle, M., Pitscheider, A., Pohlkamp, H., Scharffe, D., Specht, K., and Wiedensohler, A.: CARIBIC - Civil aircraft for global measurement of trace gases and aerosols in the tropopause region, J. Ocean. Atmos. Technol., 16, 1373-1383, doi: 10.1175/1520-0426(1999)016<1373:ccafgm>2.0.co;2, 1999.

Brown et al. - Mapping coastal optical and biogeochemical variability using an autonomous underwater vehicle and a new bio-optical inversion algorithm, 2004
https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.4319/lom.2004.2.262

Bundke, U., Pappalardo, G., Petzold, A., and Tarniewicz, J.: Report on integration across networks: common strategy and common sensors for LIDAR and aerosol extinction measurements, ENVRIplus, Brussels, Deliverable 1.4, 17 pp., 2016.

Conkright and Gregg - Comparison of global chlorophyll climatologies: In situ, CZCS, Blended in situ-CZCS and SeaWiFS, 2003
http://imars.usf.edu/~carib/Public/Refs/Gregg_IJRS.pdf

De Laat, A. T. J., Dijkstra, R., Schrijver, H., Nedelec, P., and Aben, I.: Validation of six years of SCIAMACHY carbon monoxide observations using MOZAIC CO profile measurements, Atmos. Meas. Tech., 5, 2133-2142, doi: 10.5194/amt-5-2133-2012, 2012.

De Laat, A. T. J., Aben, I., Deeter, M., Nédélec, P., Eskes, H., Attie, J. L., Ricaud, P., Abida, R., El Amraoui, L., and Landgraf, J.: Validation of nine years of MOPITT V5 NIR using MOZAIC/IAGOS measurements: biases and long-term stability, Atmos. Meas. Tech., 7, 3783-3799, doi: 10.5194/amt-7-3783-2014, 2014.

De Mazière, M., Thompson, A. M., Kurylo, M. J., Wild, J. D., Bernhard, G., Blumenstock, T., Braathen, G. O., Hannigan, J. W., Lambert, J.-C., Leblanc, T., McGee, T. J., Nedoluha, G., Petropavlovskikh, I., Seckmeyer, G., Simon, P. C., Steinbrecht, W., and Strahan, S. E.: The Network for the Detection of Atmospheric Composition Change (NDACC): history, status and perspectives, Atmos. Chem. Phys., 18, 4935-4964, https://doi.org/10.5194/acp-18-4935-2018, 2018

D'Ortenzio et al. - White Book on Oceanic Autonomous Platforms for Biogeochemical Studies: Instrumentation and Measure (PABIM), 2011
http://www.coriolis.eu.org/content/download/3150/23513/file/2009_PABIM_white_book_version1.3.pdf

Eckstein, J., Ruhnke, R., Zahn, A., Neumaier, M., Kirner, O., and Braesicke, P.: An assessment of the climatological representativeness of IAGOS-CARIBIC trace gas measurements using EMAC model simulations, Atmos. Chem. Phys., 17, 2775-2794, doi: 10.5194/acp-17-2775-2017, 2017.

ESTOC (Estimated State of Global Ocean for Climate Research)
http://www.godac.jamstec.go.jp/estoc/e/

European Commission - Maritime Spatial Planning
https://ec.europa.eu/maritimeaffairs/policy/maritime_spatial_planning_en

Falkowski et al.- Mix and match: how climate selects phytoplankton https://www.nature.com/articles/nrmicro1751, 2007

Gehlen et al. - Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, 2015
https://www.tandfonline.com/doi/full/10.1080/1755876X.2015.1022350

Haentjens et al. (2017) - Revisiting Ocean Color algorithms for chlorophyll a and particulate organic carbon in the Southern Ocean using biogeochemical floats
https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017JC012844

Heue, K. P., Brenninkmeijer, C. A. M., Wagner, T., Mies, K., Dix, B., Friess, U., Martinsson, B. G., Slemr, F., and van Velthoven, P. F. J.: Observations of the 2008 Kasatochi volcanic SO2 plume by CARIBIC aircraft DOAS and the GOME-2 satellite, Atmos. Chem. Phys., 10, 4699-4713, doi: 10.5194/acp-10-4699-2010, 2010.

Heue, K. P., Brenninkmeijer, C. A. M., Baker, A. K., Rauthe-Schoech, A., Walter, D., Wagner, T., Hormann, C., Sihler, H., Dix, B., Friess, U., Platt, U., Martinsson, B. G., van Velthoven, P. F. J., Zahn, A., and Ebinghaus, R.: SO2 and BrO observation in the plume of the Eyjafjallajokull volcano 2010: CARIBIC and GOME-2 retrievals, Atmos. Chem. Phys., 11, 2973-2989, doi: 10.5194/acp-11-2973-2011, 2011.

Hogan, R. J O Connor, E. J. Illingworth, A. 2009, Verification of cloud fraction forecasts. Quarterly Journal of the Royal Meteorological Society. 135. 1494-1511. 10.1002/qj.481
Illingworth, A. et al. CLOUDNET Continuous Evaluation of Cloud Profiles in Seven Operational Models Using Ground-Based Observations. Bulletin of the American Meteorological Society, 2007

IOCCG report 11 - Bio-Optical Sensors on Argo Floats, 2011
http://ioccg.org/wp-content/uploads/2015/10/ioccg-report-11.pdf

IUCN (International Union for Conservation of Nature)
https://www.iucn.org/theme/marine-and-polar/our-work/marine-protected-areas

Jacox et al. - The potential for improving remote primary productivity estimates through subsurface chlorophyll and irradiance measurement
https://www.sciencedirect.com/science/article/pii/S0967064513004542, 2013

Jeffrey and Humphrey - New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton, 1975
https://www.sciencedirect.com/science/article/pii/S0015379617307783

Kourafalou et al.- Coastal Ocean Forecasting: system integration and evaluation, 2015
https://www.tandfonline.com/doi/full/10.1080/1755876X.2015.1022336

Lamquin, N., Stubenrauch, C. J., Gierens, K., Burkhardt, U., and Smit, H.: A global climatology of upper-tropospheric ice supersaturation occurrence inferred from the Atmospheric Infrared Sounder calibrated by MOZAIC, Atmos. Chem. Phys., 12, 381-405, doi: 10.5194/acp-12-381-2012, 2012.

Lavigne et al. - On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea: a basin-scale and seasonal approach
https://hal.sorbonne-universite.fr/hal-01207401/document, 2015

Lavigne et al- Towards a merged satellite and in situ fluorescence ocean chlorophyll product
https://www.biogeosciences.net/9/2111/2012/bg-9-2111-2012.html, 2012

Lorenzen et al.– Determination of Chlorophyll and Pheo-Pigments: Spectrophotometric Equations, 1967

Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C., and Cook, T.: Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, An overview, J. Geophys. Res., 103, 25631-25642, doi: 10.1029/98jd00977, 1998.

Maxwell and Johnson- Chlorophyll fluorescence - a practical guide
https://academic.oup.com/jxb/article/51/345/659/652534, 2000

Mayot et al.- Physical and Biogeochemical Controls of the Phytoplankton Blooms in North Western Mediterranean Sea: A Multiplatform Approach Over a Complete Annual Cycle (2012–2013 DEWEX Experiment), 2017
https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1002/2016JC012052

Mignot et al.- From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-a concentration https://www.biogeosciences.net/8/2391/2011/, 2011

Morel and Berthon- Surface pigments, algal biomass profiles, and potential production of the euphotic Layer – Relationships reinvestigated in view of remote-sensing applications, 1989
https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.4319/lo.1989.34.8.1545

Morel and Maritorena - Bio-optical properties of oceanic waters: A reappraisal
https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2000JC000319, 2001.

Mueller et al. - Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 3, Volume 2, 2003

NASA - Coastal Zone Color Scanner Experiment (CZCS)
https://oceancolor.gsfc.nasa.gov/data/czcs/

Organelli et al. - Bio-optical anomalies in the world's oceans: An investigation on the diffuse attenuation coefficients for downward irradiance derived from Biogeochemical Argo float measurements, 2017

OSMOSIS (Ocean Surface Mixing, Ocean Sub-mesoscale Interaction Study)
https://www.bodc.ac.uk/projects/data_management/uk/osmosis/

Papagiannopoulos, N., Mona, L., Alados-Arboledas, L., Amiridis, V., Baars, H., Binietoglou, I., Bortoli, D., D'Amico, G., Giunta, A., Guerrero-Rascado, J. L., Schwarz, A., Pereira, S., Spinelli, N., Wandinger, U., Wang, X., and Pappalardo, G.: CALIPSO climatological products: evaluation and suggestions from EARLINET, Atmos. Chem. Phys., 16, 2341-2357, https://doi.org/10.5194/acp-16-2341-2016, 2016.

Pappalardo, G., et al. EARLINET correlative measurements for CALIPSO: First intercomparison results, *J. Geophys. Res.*, 115, D00H19, doi:10.1029/2009JD012147, 2010

Pasqueron de Fommervault et al. (2017) - Temporal variability of chlorophyll distribution in the Gulf of Mexico: bio-optical data from profiling floats, 2017
https://www.biogeosciences.net/14/5647/2017/bg-14-5647-2017.pdf

Perim de Faria, J., Bundke, U., Berg, M., Freedman, A., Onasch, T. B., and Petzold, A.: Airborne and laboratory studies of an IAGOS instrumentation package containing a modified CAPS particle extinction monitor, Aerosol Sci. Technol., 51, 1240-1253, doi: 10.1080/02786826.2017.1355547, 2017.

Petetin, H., Jeoffrion, M., Sauvage, B., Athier, G., Blot, R., Boulanger, D., Clark, H., Cousin, J.-M., Gheusi, F., Nedelec, P., Steinbacher, M., and Thouret, V.: Representativeness of the IAGOS airborne measurements in the lower troposphere, Elem Sci Anth., 6, 23, doi: 10.1525/elementa.280, 2018.

Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A. M., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D., Marshall, J., Nédélec, P., Smit, H. G. J., Frieß, U., Flaud, J.-M., Wahner, A., Cammas, J.-P., Volz-Thomas, A., and IAGOS-Team: Global-Scale Atmosphere Monitoring by In-Service Aircraft – Current Achievements and Future Prospects of the European Research Infrastructure IAGOS, Tellus B, 67, 28452, doi: 10.3402/tellusb.v67.28452, 2015.

Roesler et al.- Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors, 2017
https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.1002/lom3.10185,

Sauzède et al. - Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton community composition from in situ fluorescence profiles: A method based on a neural network with potential for global-scale applications, 2015
https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1002/2014JC010355

Schaeffer et al. - Mean hydrography on the continental shelf from 26 repeat glider deployments along Southeastern Australia, 2016
https://www.nature.com/articles/sdata201670#ref-link-section-24

Tanimoto, H., Zbinden, R. M., Thouret, V., and Nedelec, P.: Consistency of tropospheric ozone observations made by different platforms and techniques in the global databases, 67, doi: 10.3402/tellusb.v67.27073, 2015.

Tocquer, F., Barret, B., Mari, C., Le Flochmoen, E., Cammas, J. P., and Sauvage, B.: An upper tropospheric 'ozone river' from Africa to India during the 2008 Asian post-monsoon season, Tellus B, 67, doi: 10.3402/tellusb.v67.25350, 2015.

Uitz et al. - Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll
https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2005JC003207, 2006

UNESCO (United Nations Educational, Scientific and Cultural Organization) / IOC (Intergovernmental Oceanographic Commission) - Marine Spatial Planning Programme
http://msp.ioc-unesco.org/

Verma, S., Marshall, J., Parrington, M., Agustí-Panareda, A., Massart, S., Chipperfield, M. P., Wilson, C., and Gerbig, C.: Extending methane profiles from aircraft into the stratosphere for satellite total column validation using the ECMWF C-IFS and TOMCAT/SLIMCAT 3-D model, Atmos. Chem. Phys., 17, 6663-6678, doi: 10.5194/acp-17-6663-2017, 2017.

Wandinger, U., M. Tesche, P. Seifert, A. Ansmann, D. Müller, and D. Althausen (2010), Size matters: Influence of multiple scattering on CALIPSO light-extinction profiling in desert dust, *Geophys. Res. Lett.*, 37, L10801, doi:10.1029/2010GL042815.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, doi: 10.1038/sdata.2016.18, 2016.

Xing et al, - Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean) https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.4319/lom.2012.10.483 , 2012

Zbinden, R. M., Cammas, J. P., Thouret, V., Nédélec, P., Karcher, F., and Simon, P.: Mid-latitude tropospheric ozone columns from the MOZAIC program: climatology and interannual variability, Atmos. Chem. Phys., 6, 1053-1073, doi: 10.5194/acp-6-1053-2006, 2006.

Zbinden, R. M., Thouret, V., Ricaud, P., Carminati, F., Cammas, J. P., and Nédélec, P.: Climatology of pure tropospheric profiles and column contents of ozone and carbon monoxide using MOZAIC in the mid-northern latitudes (24° N to 50° N) from 1994 to 2009, Atmos. Chem. Phys., 13, 12363-12388, doi: 10.5194/acp-13-12363-2013, 2013.