Stakeholder Needs Review Report European Metrology Network for Climate and Ocean Observation

Version 1.0 (01/2021)

EURAMET

CLIMATE AND
OCEAN OBSERVATION

Authorship and Imprint

This document was developed by the EURAMET e.V., European Metrology Network (EMN) for Climate and Ocean Observation.

Authors: Emma Woolliams (NPL, United Kingdom), Céline Pascale (METAS. Switzerland), Paola Fisicaro (LNE, France), Nigel Fox (NPL, United Kingdom), Annarita Baldan (VSL, Netherlands), Pieter De Vis (NPL, United Kingdom), Dolores Del Campo (CEM, Spain), Miruna Dobre (SMD, Belgium), Nicole George (NPL, United Kingdom), Maria Carmen Garcia Izquierdo (CEM, Spain), Tom Gardiner (NPL, United Kingdom), Maitane Iturrate-Garcia (METAS, Switzerland), Andrea Merlone (INRIM, Italy), Rosalinda Morrone (NPL, United Kingdom), Stefan Persijn (VSL, Netherlands), Chris Rennick (NPL, United Kingdom), Stephen Robinson (NPL, United Kingdom), Marc Roche (SMD, Belgium), Steffen Seitz (PTB, Germany), Sarah Taylor (NPL, United Kingdom), Adriaan van der Veen (VSL, Netherlands), Konstantina Vasilatou (METAS, Switzerland).

Part of this work has received funding from the EMPIR programme (Project 18NET04 ‘Support for a European Metrology Network for climate and ocean observation’). The EMPIR initiative is co-funded by the European Union’s Horizon 2020 research and innovation programme and the EMPIR participating states.

Version 1.0 (01/2021)

EURAMET e.V.
Bundesallee 100
D-38116 Braunschweig
Germany

E-Mail: secretariat@euramet.org
Phone: +49 531 592 1960

Official language

The English language version of this document is the definitive version.

Copyright

ISBN 978-3-942992-60-2

Image on cover page: Contains modified Copernicus Sentinel data (2017), processed by Sinergise/ESA.
http://www.esa.int/ESA_Multimedia/Images/2018/04/Cloud-free_Europe

Further information

This is the first version of the summary report on stakeholder needs including the three ECV themes (Atmosphere, Ocean and Land), and also covering Earth Observation and other measurement techniques and synergies between themes.

No representation is made, nor warranty given that this document or the information contained in it will be suitable for any particular purpose. In no event shall EURAMET, the authors or anyone else involved in the creation of the document be liable for any damages whatsoever arising out of the use of the information contained herein. The parties using the Stakeholder Needs Review Report shall indemnify EURAMET accordingly.

For further information about this document, please contact your national contact person of the EURAMET European Metrology Network for Climate and Ocean Observation (see https://www.euramet.org/climate-and-ocean-observation).

CONTENT

Abbreviations and Acronyms

Abbreviation or Acronym	Definition / Full form	Notes / Comments
ACIX	Atmospheric Correction Intercomparison Exercise	Of radiative transfer codes, organised by CEOS
ACTRIS	European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases	
AGAGE	Advanced Global Atmospheric Gases Experiment	
ALT	Active layer thickness	Of permafrost
AMN	Atmospheric Monitoring Network	
ARD	Analysis Ready Data	For satellite data
BC	Black Carbon	
Belspo	Belgian Federal Science Policy Office	
BIPM	Bureau International des Poid et Mesures	International Bureau of Weights and Measures
BOA	Bottom of Atmosphere	Used for satellite data
C3S	European Union's Copernicus Climate Change Service	Part of Copernicus
Cal/Val	Calibration/Validation	
CAMS	Copernicus Atmosphere Monitoring Service	
CARD4L	CEOS Analysis Ready Data for Land	A form of ARD
CCL	Central Calibration Laboratory	Part of WMO-GAW
CCPR	Consultative Committee for Photometry and Radiometry	Part of CIPM
CCT	Consultative Committee for Thermometry	Part of CIPM
CCQM-GAWG	Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology - Working Group on Gas Analysis	Part of CIPM
CDR	Climate Data Record	
CDS	Climate Data Store	
CEOS	Committee on Earth Observation Satellites	
CF	Climate and Forecast	
CFC	Chlorinated Fluorinated Compound	
CFH	Cryogenic Frost-point Hygrometer	
CGLS	Copernicus Global Land Services	
CIMO	Commission for Instruments and Methods of Observation	Part of WMO (currently under reorganisation)
CIPM	International Committee for Weights and Measures	
CLD	Chemiluminescence Detection	
CMEMS	Copernicus Marine Environment Monitoring Service	Part of Copernicus
CNES	Centre National D'Etudes Spatiales	French National Centre for Space Studies

CTD	Conductivity, Temperature, Depth probes	Usually on Argo floats, for ocean observation
DI	Designated Institute	Formal institute which provides metrological services (delegated from the NMI) within the Convention of the Metre Framework
DIC	Dissolved Inorganic Carbon	
DLR	Deutsches Zentrum für Luft- und Raumfahrt	German Aerospace Centre
DQO	Data Quality Objective	Of WMO-GAW
EA	European Accreditation	
ECMWF	European Centre for Medium-Range Weather Forecasts	
ECV	Essential Climate Variable	Defined by GCOS
EEA	European Environment Agency	
EMN	European Metrology Network	Structure within EURAMET
EMPIR	European Metrology Programme for Innovation and Research	
EMRP	European Metrology Research Programme	
EO	Earth Observation	Can be used to mean all observations of the environment or can be used specifically for satellite-based observations (depends on context)
EOOS	European Ocean Observation System	
EOV	Essential Ocean Variable	Defined by GOOS, includes all ocean ECVs and additional EOVs
EQC	Evaluation and Quality Control	Programme within C3S
ESA	European Space Agency	
EUMETSAT	The European Organisation for the Exploitation of Meteorological Satellites	
EURAMET	European Association for National Metrology Institutes	
EuroGOOS	European Global Ocean Observing System	
FAPAR	Fraction of absorbed photosynthetically active radiation	One of the land ECVs
FCDR	Fundamental Climate Data Record	Term used by satellite community to describe records of Level 1 products (raw data, usually radiance values) from satellites that have sufficient quality information for climate applications

FDR	Fundamental Data Record	Similar to FCDR, but without emphasising climate applications
FIDUCEO	FIDelity and Uncertainty in Climate data records from Earth Observation	An H2020 project
FIR	Far Infrared	
FRM	Fiducial Reference Measurement	Instrumented ground network sites used for the calibration and validation of satellite observations
FRM4STS	Fiducial Reference Measurements for Surface Temperatures from Satellites	ESA project
JMA	National Meteorological Agency of Japan	
GAIA-CLIM	Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring	H2020 project
GAID	Gaps Assessment and Impacts Document	
GAW	Global Atmosphere Watch	Programme of WMO
GCOS	Global Climate Observing System	co-sponsored by WMO, IOC-UNESCO, UNEP and the ISC
GEO	Intergovernmental Group on Earth Observations	
GEOSS	Global Earth Observation System of Systems	Programme of GEO
GEWEX	Global Energy and Water Cycle Experiment	
GHG	Greenhouse Gas	
GOA-ON	Global Ocean Acidification Observing Network	
GOOS	Global Ocean Observation System	
GOS	Global Observing System	
GRUAN	GCOS Reference Upper-Air Network	
GSICS	Global Space-based InterCalibration System	
GTN-P	The Global Terrestrial Network for Permafrost	
GUM	Guide to the Expression of Uncertainty in Measurement	
HCFC	Chlorinated Fluorinated Hydrocarbon	
HFC	HydroFluoroCarbon	
IAPSO	International Association for the Physical Sciences of the Oceans	
ICOS	Integrated Carbon Observation System	
ISC	International Science Council	
ILAC	International Laboratory Accreditation Cooperation	
ILC	Interlaboratory comparison	
IMOP	Instruments and Methods of Observation Programme	

IMS	International Monitoring System	
INSPIRE	Infrastructure for Spatial Information in Europe	
IOC	Intergovernmental Oceanographic Commission	Part of UNESCO
IODE	International Oceanographic Data and Information Exchange	
IOM		Report series title for reports of WMO Instruments and Methods of Observation programme
IP	Intellectual Property	
IPA	International Permafrost Association	
IPCC	Intergovernmental Panel on Climate Change	
IR	Infrared	
ISMN	International Soil Moisture Network	
ISO/IEC	International Organization for Standardization/International Electrotechnical Commission	
JRC	Joint Research Centre	Of the European Union
JRP	Joint Research Project	Of EMPIR and EMRP programmes
LiDAR	Light Detection and Ranging	
MBES	Multi-beam Echo-sounder	
MCMC	Markov Chain Monte Carlo	
MIR	Mid Infrared	
MOU	Memorandum of Understanding	
MRA	Mutual Recognition Arrangement	Formal part of metrology
NDIR	Non-dispersive infrared	
NMHC	Non-Methane Hydrocarbon	
NMI	National Metrology Institute	Formal institute representing metrology within the Convention of the Metre Framework
NO_x	Nitrogen Oxides	
NWP	Numerical Weather Prediction	
OA	Ocean Acidification	
OBPS	Ocean Best Practices System	
OE	Optimal Estimation	
pCO2	Partial pressure of carbon dioxide	
PM	Particulate Matter	
PTR-MS	Proton Transfer Reaction – Mass Spectrometry	
QA / QC	Quality Assurance / Quality Control	
QA4EO	Quality Assurance Framework for Earth Observation	
QA4SM	Quality Assurance Framework for Soil Moisture	
RCC	Regional Calibration Centre	Of WMO-GAW

ROOS	Regional Ocean Observing System	Part of GOOS, EuroGOOS is an example
RTM	Radon Tracer Method	Note potential confusion as the same acronym is used for bothe. See sections 3.1.1.2 and 6.5.
RTM	Radiative Transfer Model	
SDG	Sustainable Development Goal	United Nations framework
SI	International System of Units	
SO_2	Sulfur Dioxide	
SPM	Suspended Particulate Matter	
SSW	Standard Seawater	Standardised by IAPSO
SWIR	Short-wave infrared	
TA	Total Alkalinity	
TOA	Top of Atmosphere	
UKSA	United Kingdom Space Agency	
UN	United Nations	
UNEP	United Nations Environment Programme	
UNESCO	United Nations Educational, Scientific and Cultural Organization	
USGS	United States Geological Survey	
UTLS	Upper Troposphere/Lower Stratosphere	
UV	Ultraviolet	
VIM	International Vocabulary of Metrology	
VIS	Visible	
VOC	Volatile Organic Compound	
VPDB	Vienna Pee Dee Belemnite	Scale reference for $^{13}C/^{12}C$ ratio
WHP	World Hydrographic Program	
WISG	World Infrared Standard Group	Maintained as part of the World radiometric reference
WMO	World Meteorological Organization	
WMO-CCl	World Meteorological Organization Commission for Climatology	
WMO-CIMO	WMO Commission for Instruments and Methods of Observation	
WOCE	World Ocean Circulation Experiment	

Executive Summary

This Executive Summary is a synthesis of the Stakeholder Needs Report of the European Metrology Network for Climate and Ocean Observation (here “the EMN”). During 2020, the EMN carried out a review to identify and prioritise the ways in which metrology can support climate and ocean observation. The review report highlights the areas that we found that require urgent further collaborative metrological research between metrology institutes and experts in those observation systems and their applications.

This report uses the terms “climate and ocean observation” and “metrology community” in specific ways:

- The EMN has two related, but separate, themes – the “climate” theme relates to metrology for supporting the making and use of observations of essential climate variables (ECVs) in all three domains (atmosphere, ocean and land). The “ocean” theme relates to metrology for supporting observations of essential ocean variables (EOVs). These EOVs include all the ECVs in the ocean theme along with additional variables to cover a broader range of applications, including those needed to understand ocean biodiversity and to support the sustainable use of the ocean for cultural, social and economic benefit and as a food supply. Thus “climate and ocean observation” is used to describe both themes of the EMN and not to suggest any particular emphasis on the ocean within the “climate” theme.

- We recognise that there is a multidisciplinary community making and using climate and ocean observations and that people within those communities rightly also consider themselves “metrologists”. Likewise, many people working in formal metrological institutes are already active in committees, consortia and networks involved in climate and ocean observation. Here, however, we use “the metrology community” and “metrologists” to only describe those working in institutes that are formally recognised by the Metre Convention as national metrology institutes (NMIs) and designated institutes (DIs). Within Europe, such institutes are members of the European Association for National Metrology Institutes (EURAMET) and eligible to be formal members of the EMN.

Metrology in Climate and Ocean Observation

To mitigate and adapt to climate change, decision makers in governments, industry and non-governmental organisations need access to high quality information about the historical, current and future state of the climate system and on the current anthropogenic greenhouse gas (GHG) emissions, water use and land use changes. Such information relies on direct observations of the state of the environment. Near-real time observations are used in climate and ocean data services and in numerical weather prediction (NWP) forecasts for a wide variety of societal applications. Longer-duration historical climate data records of essential climate variables (ECVs) are assimilated into climate reanalyses (which provide detailed information on the state of the historical climate). These long-term observationally-derived datasets are also key to validating and improving climate models and therefore help to improve our understanding of the physics, chemistry and biology of the Earth system. The ECVs are defined by the Global Climate Observing System (GCOS).

Figure 1 shows how climate observations fit into the climate decision making process. The outer loop (blue arrows) shows how individual observations (which come from both in situ and remote sensing methods) of the climate system are processed into climate data records (CDRs) of ECVs, which in turn are used to tune and validate climate models. The climate models inform, and are informed by, integrated assessment models, models used by economists and social scientists to understand and predict changes in anthropogenic GHG emissions and land use. These integrated assessment models inform policy, which leads to changes in emissions and land use, which affect the climate (on decadal timescales) and the changed climate can be monitored by observations. The fast feedback loop (gold) provides near real-time (timescale 2-3 years) observations of GHG sources (emissions) and sinks (e.g. forests). These observations can support national inventories and thus the 'Global Stocktake' reports for the Paris Agreement. The changes in anthropogenic GHG emissions and land use can be observed directly, providing feedback to policy on timescales of a few years.

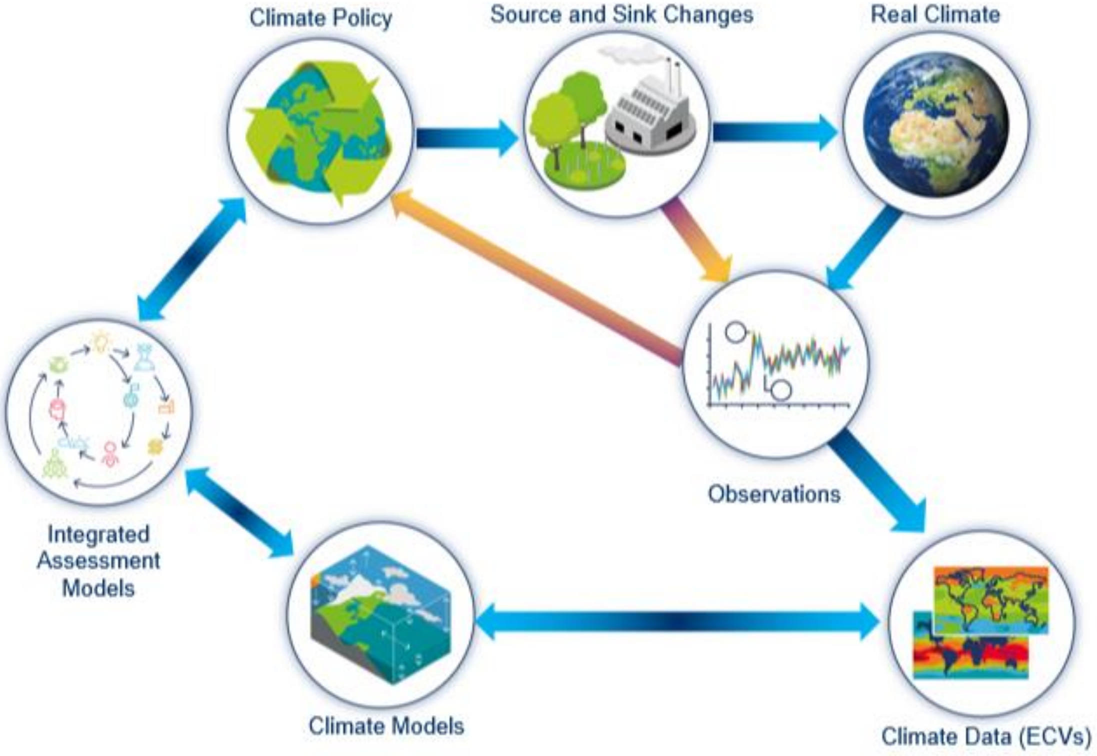


Figure 1 The role of climate observations within the decision cycle – using greenhouse gas emissions as an example where society can have a direct impact.

Some of the biggest decisions that humanity must make are based on climate models and the integrated assessment models they inform. Climate policy in many countries (and indeed many commercial enterprises) is now to achieve 'net zero' emissions by the middle of the century, or earlier. This is highly ambitious and requires fresh thinking and reliable data to support the required policy changes. It is therefore essential to have quality assurance (QA) of the entire value chain shown in Figure 1. Such QA requires reliable observations that are stable long term, linked to a common reference (ideally the international system of units, SI) and have robustly-determined uncertainties associated with them. It also requires that the models used in the data analysis are themselves quality assured and have associated uncertainties. Note that models are used to interpret raw observations, and to process such observations into ECVs, as well as in climate prediction and reanalysis.

Metrology, the science of measurement, can contribute to both the QA of observations and, through data scientists working in metrology institutes, to the QA of data processing and modelling. Metrology has ensured that the SI units have been stable for nearly 150 years, are consistent worldwide and are coherent, through the key principles:

1. Metrological traceability: Linking measurement results to a commonly-agreed primary reference (the SI);
2. Comparison between independent measurement approaches: both exploratory scientific comparisons and formal comparisons that support the formal “Mutual Recognition Arrangement” (MRA) between metrology institutes, and
3. Uncertainty analysis following the principles of the Guide to the Expression of Uncertainty in Measurement (GUM).

These key principles of metrology can, should and are being applied to the observations of our environment and our climate, albeit with necessary interpretation and adjustment, to provide QA to the observations that society relies on.

Review of Stakeholder Needs

The stakeholders of the EMN are those who make or use climate and ocean observations. These include those who manufacture and calibrate sensors deployed in in situ networks and satellite sensors, those who operate observational networks and satellite systems, those who process raw data from current and historical networks and systems to give “fundamental data records” of raw data or derived “climate/thematic data records” of ECVs and EOVs, and those who use such observational records in climate or oceanographic models or to provide societal benefit or commercial services derived from these records. Most of these stakeholder communities are organised through international organisations and their committee structures and the EMN has sought to interface directly with those existing organisations.

During 2020, the EMN collated needs from stakeholder communities using an online survey, four webinar workshops and by participating in stakeholder committees. The EMN also reviewed summaries of historical workshops, scientific literature, and stakeholder community strategies and implementation plans to identify and prioritise the areas where metrology can most meaningfully contribute.

The stakeholder needs report describes the outcome of this review. After an introduction, it has five main sections, describing the metrological needs for:

- Atmosphere observation from ground-based instruments (the GCOS Atmosphere ECVs which are measured with in situ methods)
- Ocean observation from in situ instruments (the GCOS Ocean ECVs and additional Global Ocean Observation System (GOOS) EOVs)
- Land observation from in situ instruments (the GCOS Land ECVs which are measured with in situ methods)
- Remote sensing of ECVs and EOVs (by satellite-based, aircraft-based and ground-based remote sensing methods)
- General observations (a section that considers common needs across multiple themes)

In each section of the report there is a discussion of the identified needs, including quotes from our survey and references to the reports in which the need was identified. Each section ends with a table of priority requirements for that theme. Those tables are replicated in Appendix A of the Stakeholder Needs Report and in Section 5 of the stand-alone version of the Executive Summary.

Metrology Challenges for Climate and Ocean Observation

The traditional role of metrology in providing SI-traceability to observational systems, through reference standards, materials, instrumentation and associated calibration, remains important.

For some ECVs, SI traceability is already routinely provided through NMIs or DIs and there are ongoing requirements for operational services of this nature, iterative improvements and improved access to standards for field measurements. This is particularly the case for the provision of traceability to remote locations and/or for the low concentration levels of a given substance in the atmosphere / ocean. For other ECVs, the community has existing community standards that, while not SI-traceable, are considered at present to be fit for purpose, although many of these communities need support to assess the uncertainty and traceability to these standards. However, for several ECVs and observational methods, significantly improved and even new reference and working standards, materials and instrumentation are required – spanning measurements made in NMI laboratories, calibration laboratories, in the field and in space.

In a few cases, particularly for water vapour concentration in the atmosphere (humidity) and ocean pH, the definition of the measurand itself is ambiguous, with multiple possible definitions that depend on the method for taking the observation. Here, metrologists can work with the observational communities to support a standardisation of the definition and to provide comparisons and conversion factors between different methods.

Metrology institutes are already involved in the establishment of primary reference networks, including, for example, the GCOS Reference Upper Air Network, GRUAN, the planned establishment of a surface reference network, and the establishment of fiducial reference measurement (FRM) sites for satellite post-launch calibration and validation. Such reference networks are needed because many observations were not originally taken for climate purposes, and these provide the long-term stability and absolute accuracy required for a CDR. It is essential that metrology institutes continue to support the development of reference networks including in the provision of traceability and in establishing uncertainty analyses for the observed quantity in the environmental conditions.

There are several identified needs relating to running comparisons for observation communities. Such comparisons are needed at all the different levels: interlaboratory calibrations of calibration facilities, in field comparisons of observational instruments, comparisons of satellite data products over reference sites and comparisons of derived ECV products, including the algorithms used to process data (e.g. atmospheric correction of satellite data) and to generate ECVs.

More generally, there is a role for metrologists in supporting standardisation through definitions (as described above), through good practice guides – in some cases formal international standards – and through developing quality assurance and quality control (QA/QC) frameworks for observational practices and data provision. QA/QC frameworks can support requirements for traceability to the SI or a community-agreed reference and support the need for rigorous uncertainty assessment. Such frameworks are needed both for current observational networks and satellite sensors and for “data rescue” of historical data. Related to this there is a need to support the correct use of metrological vocabulary (e.g. uncertainty/error, traceability) in both vocabularies (written for scientists) and ontologists (written for computer databases) and to provide training in metrological techniques (uncertainty analysis and practical techniques).

There are also requirements for metrological data science. These requirements include the development of methods to propagate uncertainties and error covariance structures through data assimilation, reanalyses and models. They include developing metrological methods for assessing the uncertainties associated with complex processing chains, including neural networks and classification processes. They include considering the uncertainties inherent in models. Data science techniques are also needed to determine outliers in large networks of autonomous sensors, particularly for low-cost sensors, and to understand scaling and representativeness as individual observations on different scales are compared.

About the European Metrology Network for Climate and Ocean Observation

The European Association for National Metrology Institutes (EURAMET) created European Metrology Networks (EMNs) in 2018 to support its vision to ensure Europe has world-leading metrological capability based on high-quality scientific research and an effective and inclusive infrastructure that meets the rapidly advancing needs of end users. The first six EMNs were established in early 2019 with a remit to analyse the European and global metrology needs and to seek to address those needs in a coordinated manner through the National Metrology Institutes (NMIs) and Designated Institutes (DIs) of Europe. EMN members will formulate common metrology strategies, including scientific research, infrastructure development, knowledge transfer and service offerings. The EMNs will provide a single point of contact for information, underpinning regulation and standardisation, and for promoting best metrological practice.

One of the first EMNs to be established was the EMN for Climate and Ocean Observation. The EMN has a scope that covers metrological support for in situ, ground-based and remote sensing observations of atmosphere, land and ocean Essential Climate Variables (ECVs) for climate observations and also to support the broader economic and ecological applications of Essential Ocean Variables (EOVs) observations. It is the European contribution to a global effort to further enhance metrological best practice into such observations through targeted research efforts.

1. INTRODUCTION

1.1. Metrology in Climate and Ocean Observation

In 2015, responding to humanity's responsibility to limit and reverse our environmental damage while improving human health and quality of life, the United Nations (UN) ratified the "Sustainable Development Goals" (SDGs) [1]. The SDGs define humanity's joint hope for a future lived within ecological boundaries that has greater social equality, improved human health and improved quality of life. The SDGs are defined by 17 goals. Amongst these, goal 13 states: "Take urgent action to combat climate change and its impacts" and SDG 14 states: "Conserve and use the oceans, seas and marine resources for sustainable development".

We are also at the start of an information revolution. The rapid increase in computational resources and mathematical tools for handling "big data", combined with the ever-increasing deployment of operational observational systems and the operational provision of environmental information services, such as the European Copernicus Services, have made information on the current (live), historical and predicted states of the environment available to everyone, at any time and in any place.

However, with an ever-increasing quantity of environmental information and observations, it is essential that these are underpinned by robust quality assurance (QA). To identify a small climate trend from an observational record that is also sensitive to changes in weather, to seasonal effects, and to geophysical processes, it is essential that observations are stable over multiple decades, while still allowing for changes in the observation instrumentation and operational procedures. The resultant data should also be traceable to a common reference, with well-understood uncertainty analysis, so that observations are interoperable and coherent; in other words, measurements by different organisations, different instruments and different techniques should be able to be meaningfully combined and compared.

It is here that metrology, the science of measurement, can contribute. Metrology has ensured that the SI units have been stable for nearly 150 years, are consistent worldwide and are coherent. Metrology has provided this consistency through the key principles:

1. Metrological traceability ([2]): Linking measurement results to a commonly-agreed primary reference (usually the international system of units, SI);
2. Comparison between independent measurement approaches: Both exploratory scientific comparisons and formal comparisons that support the formal Mutual Recognition Arrangement (MRA) [3], and,
3. Uncertainty analysis following the principles of the (GUM) [4].

These key principles of metrology can, should and are being applied to the observations of our environment and our climate, albeit with necessary interpretation and adjustment, to provide QA to the observations that society relies on.

1.1 Purpose and Scope of Document

This document has been prepared by the members of the European Metrology Network for Climate and Ocean Observation (hereafter, "the EMN"). This network has two related, but separate themes. The "climate" theme relates to metrology for supporting the making and use of observations of essential climate variables (ECVs; see Section 2.2) in all three domains (atmosphere, ocean and land). These are collected predominantly to improve our understanding of the Earth's climate, although many have other additional applications too. The "ocean" theme relates to metrology for supporting observations of essential ocean variables (EOVs; see Section 2.3). These EOVs include all the ECVs in the ocean theme along with additional variables and cover a broader range of applications, including those needed to understand ocean biodiversity and to support the sustainable use of the ocean for cultural, social and economic benefit and as a food supply. Because there are many commonalities in how metrology can engage for these two themes, our network covers both.

This document identifies and prioritises the ways in which metrology can support and improve climate and ocean observation, monitoring and alert systems. It highlights the areas that require further collaborative metrological research between metrology institutes and experts in those observation systems and their applications.

1.2 What We Mean by “the Metrology Community”

We recognise the considerable existing expertise in metrological methods in the scientific, commercial and operational communities that make and use observations, and the ongoing and growing existing collaboration between those communities and metrology institutes. In this report we have, as short-hand, talked of “the metrology community” meaning the institutes that are formally recognised by the Metre Convention – the national metrology institutes (NMIs) and associated designated institutes (DIs). This does not, in any way, mean that we underestimate or undervalue the work of professional metrologists working in organisations that make climate and/or ocean observations. Nor do we underestimate or undervalue the existing activity of people at formal metrology institutes who are already part of (and in some cases leading) committees, consortia and networks involved in climate and ocean observation. Climate and ocean observation is a multidisciplinary activity and metrology is one of those disciplines. This report, however, is to inform the formal NMIs and DIs, about what they can do more of, to increase those collaborations and participation.

1.3 Methodology for Identifying Stakeholder Needs

The conclusions presented in this report are based on the results of surveys and workshops organised by the EMN for Climate and Ocean Observation and on reports of such needs from workshops that others have organised in the last decade.

The EMN carried out a survey from November 2019 to February 2020. The survey was organised in six separate versions aimed, respectively at:

- Metrology aspects for instrumentation in climate and ocean observation
- Metrology aspects for satellite sensors in climate and ocean observation
- Metrology aspects of ECVs and EOVs records in climate and ocean observation
- Metrology aspects for information services built on ECV/EOV records
- Metrology aspects for data centre services for climate and ocean observation
- Metrology aspects for those commissioning and funding ECV/EOV services

We received 55 replies – 54 of which were for one of the first three surveys and one for the final survey. The other two surveys had no responses.

The results of the survey were analysed. Quantitative data is given in Appendix B. Written responses were all considered carefully and quotations from these are given in the relevant sections below.

After the survey, we held four webinars on the 12 and 13 February 2020. These were attended by approximately 100 participants. Each webinar had two invited speakers and a general discussion about metrological involvement in that area. The webinars were on:

- Ocean observations from in situ sensor networks
- Atmosphere observations from in situ sensor networks
- Satellite-based Earth observations
- Measuring and monitoring ECVs and EOVs

Key points from those discussions are given in the relevant sections below.

We also reviewed the reports from several key workshops that considered metrological needs in observations and read strategy documents from key organisations such as the World Metrological Organisation (WMO), the European Ocean Observation System (EOOS) and the Committee on Earth Observation Satellites (CEOS).

1.4 This Document

This document describes how metrology can support both the process of making climate and ocean observations and those who use such observations for scientific and societal benefit.

- Section 3 introduces how observations are made and used, and the key organisations involved in such observations.
- Sections 4 to 6 consider instrument networks used to measure atmosphere, ocean and land ECVs respectively and Section 7 considers remote sensing methods (predominantly, but not exclusively, from satellites).
- Section 8 looks at cross-cutting metrological requirements that cut across all these themes.
- Section 9 provides a summary of core concepts and describes the next steps.

2 OBSERVING THE CLIMATE AND THE OCEAN

2.1 The Role of Observations in Decision Making

The 2015 Paris Agreement seeks to limit the rise in temperature of the Earth to less than 2 °C, with a target of less than 1.5 °C above pre-industrial levels in order to reduce the risks and impacts of climate change. To achieve this, many countries have policies to achieve 'net zero' emissions by the middle of the century. This extremely challenging goal requires immediate, sustained and significant annual decreases in GHG emissions and a simultaneous increase in carbon sinks (in soils, forests and ocean phytoplankton). Implementing policy to achieve these highly ambitious targets requires fresh thinking and reliable data to support it. The UN SDG framework balances this need (and related urgent challenges caused by, e.g., pollution, habitat and biodiversity loss and freshwater reduction) against the human need to "leave no one behind" and to allow fair, and sustainable, economic and social development across the whole world.

The ocean is an example of the challenging balance between human and environmental needs. More than 2.4 billion people live within 100 km of a coast and many of these depend on the oceans for food security, tourism, transportation, heritage and climate regulation. Many are also vulnerable to sea level rise and from the increased storms generated by a warming ocean and atmosphere. We have over-exploited the seas, with fish stocks reducing, mineral resources over-extracted, plastics found even at the deepest parts of the ocean, and with coral reefs and other unique habitats dying from increasing temperatures and acidity.

Governments, international organisations, businesses and charities, need to make complex, inter-related decisions to meet the ambitious targets laid out in the UN SDGs. To make such decisions, they need information about the historic, present and future state of the climate. Figure 2 shows how climate observations fit into the climate decision making process. A similar diagram could be presented for other key Earth cycles, or other applications of climate records of land, ocean or atmosphere for decision making.

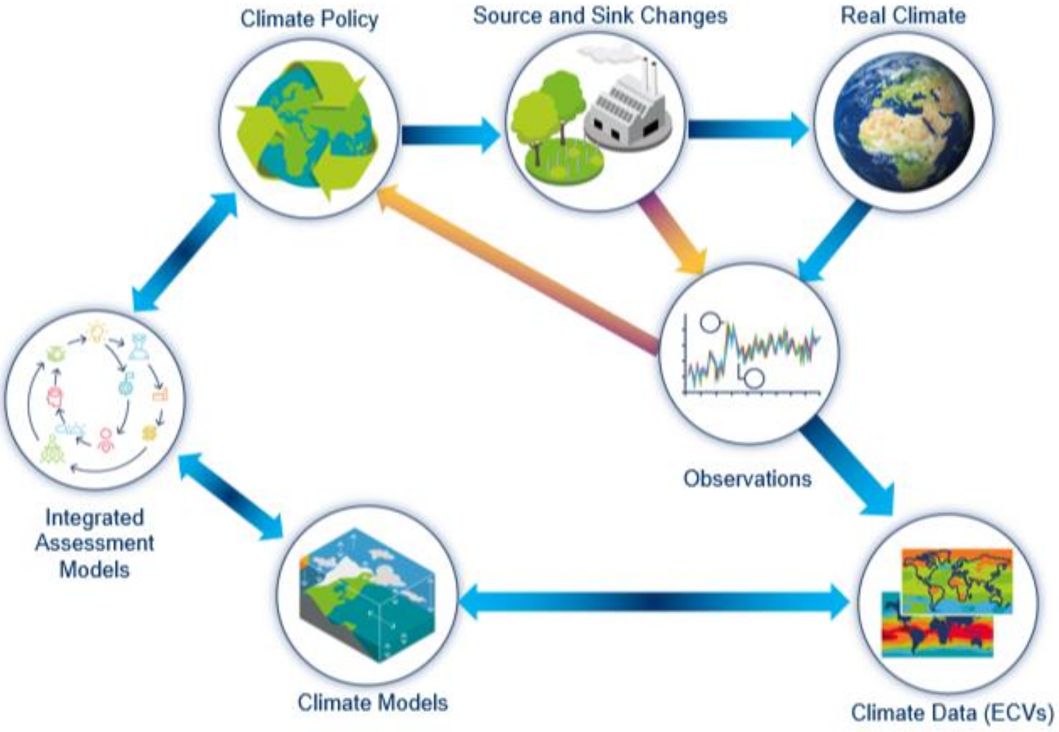


Figure 2 The role of climate observations within the decision cycle

The top line of Figure 2 shows how policy impacts the climate. Policy makers are now driving for net zero GHG emissions by mid-century with substantial progress in the next decade. This policy is implemented in changes to all sectors of society, creating reductions in GHG emissions (sources) and increases in GHG absorption by vegetation and the oceans (sinks). Reductions in sources and increases in sinks will alter the real climate of the Earth, bringing it back towards an equilibrium, ideally within 1.5 °C of pre-industrial levels. This is the process that society is relying on to prevent catastrophic climate breakdown. Observations provide essential feedback loops to this process to ensure that policy implementers are doing what they say they are doing (short feedback loop, shown with gold arrows in Figure 2) and that the Earth system is changing as expected (long feedback loop, shown with blue arrows in Figure 2). Society is making huge decisions with significant human, social, environmental and economic implications; it is essential that these feedback loops are rapid and reliable.

In the short feedback loop (gold arrows), direct observations of GHG emissions are used to test if and how these policies have been implemented by monitoring emissions and land use change, as well as monitoring natural sources and sinks of GHGs (e.g. permafrost thawing to release methane, algae in the sea absorbing CO_2, and the health of the rainforests) over timescales of 2-3 years. This feedback loop provides near-immediate and ideally global information, collated in national inventories and ‘Global Stocktake’ reports for the Paris Agreement. At present, national inventories tend to be produced from socioeconomic data, rather than from direct observations. There are significant opportunities to improve the robustness of this feedback loop, especially to support attribution studies for GHG emissions and to improve the use of satellite observations in land use and land cover change (LULCC) assessments.

The slower feedback loop (blue arrows) covers the full value chain and shows, through observations and models, how the real climate is responding to these changes. This feedback is provided by the direct observation of the state of the climate. These continued observations become part of multi-decadal climate data records (CDRs) of ECVs (see Section 2.2) and provide information on how the Earth's climate is changing. CDRs of ECVs are used by climate scientists to validate and test climate models through "hindcasts" – climate models that are started at early dates so they can be validated through comparison to CDRs. CDRs of ECVs are also used by scientists to explore the physics, chemistry and biology of specific Earth system processes to enhance scientific understanding that can improve those climate models. CDRs of ECVs are also provided as "climate data services" – for example through the European Union's Copernicus Climate Change Service. CDRs of ECVs also provide the observational data that is assimilated into climate reanalyses which provide detailed information on the historical state of the climate.

Climate models predict the changes in the Earth system; but in the Anthropocene, such changes are dominated by the economic decisions, and industrial and social actions of humanity. Climate models are therefore fed into "Integrated Assessment Models", which are developed and used by economists and social scientists to understand 'scenarios' of human activity and these are informed by and inform the climate models. The outputs of climate models and integrated assessment models are summarised in reports for policy makers (e.g. [5]). In most cases, politicians, industrialists, risk analysts, companies, charities, diplomats, and other decision makers use these outputs as a basis for their decision making).

The emissions and sinks feedback loop (gold arrows) is faster because the changes are obvious on short timescales (a few years) and because the observational data are immediately usable by policy makers, economists and in auditing. The climate science feedback loop (blue arrows) is slower because observing a climate trend in data that are noisy due to natural variability (e.g. weather, seasonal effects and effects such as the El Niño) takes multiple decades. Furthermore, such observational data are used by climate scientists and require interpretation by experts (climate scientists and then integrated assessment modellers) before they can inform policy. Nevertheless, both feedback loops require robust, quality-assured data with low uncertainties – for the emissions and sinks feedback loop because of the social, economic and environmental decisions that are directly based on those observations, and for the climate science feedback loop because observations with smaller, more robust, uncertainties significantly reduce the timescales needed to observe a climate trend.

Metrology can support the processes that ensure the QA of the observations that form the basis of both the long-term and short-term feedback loops and provide independent but robust guidance on the degree of confidence that should be attached to the resultant information. This role is explored further in this document. There is also a role in providing data science to support the overall process. That is not fully considered here but is discussed in Section 7.9.

2.2 The Global Climate Observing System and Essential Climate Variables

As discussed in the previous section, political, economic and social decisions require reliable forecasts from comprehensive, trustworthy climate models and a robust, integrated global climate observing system. Article 7, point 7c of the Paris Climate Agreement [6] states that:

> “The Parties should strengthen their cooperation […], including with regard to: […] strengthening scientific knowledge on climate, including research, systematic observation of the climate system and early warning systems, in a manner that informs climate services and supports decision-making.”

The Global Climate Observing System (GCOS) is a programme, co-sponsored by the WMO, the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO), the UN Environment Programme and the International Council of Science, which defines itself [7] as working “towards a world where climate observations are accurate and sustained, and access to climate data is free and open”. The GCOS maintains definitions of, and requirements for, ECVs [8]. ECVs have been identified and specified to be considered as the core observations that should be made systematically, globally and robustly to observe the Earth’s changing climate. These ECVs are used directly for a wide range of climate and commercial applications and are also used to validate and constrain climate models. At present there are 54 defined ECVs, associated with three domains: atmospheric, oceanic and terrestrial. Each ECV has a set of observational requirements that cover the spatial and temporal resolution of the observation and the uncertainty required both for individual measurements and for trend observations (the latter called the “stability”). Many of these requirements are difficult to achieve in a laboratory, let alone in the environment with observations over decades. Around half can only be effectively observed from space.

Furthermore, in most cases the ECVs are bio/geophysical in nature and obtained not through a direct observation, but through physical observables combined in a processing algorithm that often involves bio and/or geophysical models. Thus the “observation" becomes a combination of measurement and modelling and often involves multiple processing steps in a complex workflow. Such workflows are particularly common for space-based observations where corrections are required to account for the atmosphere in land/ocean observations and the surface in atmospheric observations. Further processing is required in all cases to create CDRs – long-term records of ECVs that must be adjusted to account for potential instrument instability and instrument changes, as well as sampling issues.

The GCOS website provides information about and locations of the observational records of ECVs that are openly accessible. These records may come from in situ observations, on site or remote sensing and include individual measurements and products that create (gap-filled) global information by merging observations with Earth system models, e.g. through meteorological reanalysis (data assimilation) or global/regional synthesis products.

Fourteen of the 17 atmospheric ECVs, six of the 18 oceanic ECVs and eleven of the 16 terrestrial ECVs require a significant, and in many cases sole, contribution from satellite Earth observation (EO). The others are observed primarily by in situ observational networks.

2.3 Ocean Observations

The oceans cover 71 % of the Earth’s surface and have absorbed almost 93 % of the enhanced anthropogenic greenhouse warming so far [9]; this has had a significant impact on the oceans ability to serve society economically. Oceans are a crucial source of food, water, energy and minerals for human life, and are a medium for transport (90 % of goods are shipped by sea), recreation and commerce. The value of marine activities globally is about 5 % of the global GDP, expecting to reach around US$3 trillion by 2030 through sustainable growth (the value following an unsustainable scenario is smaller) [10]. The European Union's Integrated Maritime Policy focuses on 'Blue Growth' – harnessing the potential of Europe's oceans, seas and coasts to stimulate economic development within the environmental boundaries of the ocean ecosystems that sustain that growth.

To balance environmental, social and commercial concerns, the Global Ocean Observing System (GOOS), a programme executed by the IOC of UNESCO, has established a framework for ocean observing, centred around EOVs. The EOVs include all the ECVs in the oceanic domain and additional variables that relate to ecosystem, disaster-warning and commercial observational requirements.

2.4 International and European Observation Programmes

In addition to a wide range of national and intergovernmental initiatives, Europe has prioritised with major long-term investment the provision of environmental information services through the creation of Copernicus [11] – the European Union's Earth Observation (EO) Programme – which provides information from satellite and in situ observing systems to its users. Copernicus is coordinated and managed by the European Commission and implemented in partnership with the Member States, the European Space Agency (ESA), the European Organisation for the Exploitation of meteorological Satellites (EUMETSAT), the European Centre for Medium-range Weather Forecasts (ECMWF), EU Agencies and Mercator Océan.

Copernicus includes data from its own set of dedicated operational satellites (the Sentinels) along with data from other commercial and public satellites and in situ networks. The Copernicus services transform this raw data into information by processing and analysing the data and making it available operationally, freely and openly accessible. Data are provided through six thematic streams: atmosphere; marine; land; climate change; security; and, emergency. The Copernicus Climate Change Service (C3S) provides observations of many of the ECVs in all three GCOS domains, while the Copernicus Marine Environment Monitoring Service (CMEMS) provides several of the EOVs. The Copernicus Atmosphere Monitoring Service (CAMS) covers both air quality measurements and climate-relevant ECV records and services and the Copernicus Global Land Services (CGLS) provides terrestrial products and services.

Through Copernicus and other observational monitoring programmes, Europe provides a significant contribution to the GCOS and the GOOS. In the ocean domain, the new (established 2016) EOOS is working to create a pan-European integrated ocean observing capacity. In the atmosphere, land and satellite domains, the European contribution is through global programmes such as the WMO's Global Atmosphere Watch (GAW) or the Intergovernmental Group on Earth Observations (GEO) Global Earth Observation System of Systems (GEOSS) and the CEOS.

3 ATMOSPHERE

Historically, key meteorological parameters − air temperature, humidity, pressure, precipitation, wind speed and direction – have been continuously monitored in a systematic way mainly for weather forecasting applications. However, the long record of global air temperature was identified as fundamental for evidencing the current global warming. Therefore, these meteorological parameters and similar ones are now also monitored for climate research activities and these parameters are defined as GCOS ECVs. These physical atmospheric parameters are measured both at the surface and in the upper atmosphere.

The WMO coordinates these observations through its own Global Observing System (GOS) programme and through co-sponsorship of GCOS alongside IOC-UNESCO, UN Environment and the International Science Council. The WMO promotes standardisation, develops guides on instruments and methods of observation, runs comparisons and provides training. The WMO has a formal participation in the metrology community, having signed the CIPM MRA in April 2010, and WMO experts participate in CIPM working groups. WMO experts also participate in EURAMET's research council and metrologists from Europe's NMIs are members and chairs of WMO committees.

Anthropogenic activities affect the atmosphere by changing its composition, which, in turn, will influence the atmospheric radiation budget and drive climate change. Thus, in order to improve our understanding of climate change and subsequent environmental impacts, global monitoring of the atmospheric composition is fundamental. The WMO's GAW Programme was established in 1989 in recognition of the need for atmospheric global observations. Besides systematic, global observations of the chemical composition of the atmosphere, the GAW provides analysed observation data to understand atmospheric changes and forecast future trends from these observations. GAW comprises observations of monitoring station networks, along with supplementary observations from satellites and aircraft. World data centres receive the observation data, process them and output GAW products – GHG and ozone bulletins, as well as global data.

The GAW Implementation Plan for 2016-2023 [12] describes the QA system in place to provide traceability and comparability to the observational networks. This system is based around the so-called Data Quality Objectives (DQOs) and the use of Central Calibration Laboratories (CCLs) and Regional Calibration Centres (RCCs) providing traceability through a hierarchical chain. The implementation plan describes a need to create consistency in methods and procedures, such as traceability, harmonised guidelines for the operation and calibration of instruments, internationally accepted methods of uncertainty analysis and metrological terminology.

At present, and building on the WMO-CIPM collaboration formalised in 2010, NMIs or DIs act as CCLs for four (surface O_3 (ozone), VOCs (Volatile Organic Compounds), NO_x (nitrogen oxides) and solar radiation) of the twenty variables being monitored by the GAW, while five of these variables have no CCL (chlorinated fluorinated compounds (CFCs), chlorinated fluorinated hydrocarbons (HCFCs) and hydrofluorocarbons (HFCs), SO_2 (sulfur dioxide), ultraviolet (UV) radiation, aerosol physical properties and aerosol chemical properties) [12].

The need for improvements with respect to the metrological traceability of the measurement results and QA is expressed by the participants of our survey. The DQOs are often not met because of lack of suitable reference materials, suitable reference data over the relevant pressure and temperature ranges or instrumental transfer standard issues. The extent to which the current practice meets emerging needs depends on the gas species measured. For CH_4

(methane), non-methane hydrocarbons (NMHCs) and some VOCs, the participants felt that traceability and QA needs are fully met. However, for other species, such as H_2O (water), NO_x, oxygenated VOCs, N_2O, black carbon and isotopes ratios, the DQO are not or only partially met. Typical issues are the chemical reactivity, short temporal stability of compounds, very low ambient concentrations or difficulties in preparation.

The main challenges are described in the following subsections. Note that this section only considers the in situ and ground-based observations of atmospheric ECVs. Where observations of radiation or atmospheric properties and composition are made using remote sensing techniques (predominantly, but not exclusively satellites) and are not traceable to a reference material or gas, these are discussed separately in Section 6.

3.1 Atmospheric Composition

3.1.1 Greenhouse Gases

GHGs, such as CO, CH_4, N_2O as well as halogenated compounds (SF_6, CFCs, HCFCs or HFCs), have a direct impact on the global temperature because of their warming potential. Measuring their atmospheric concentrations is pivotal to understanding the influence of human activity on climate. Therefore, there are many national and international legislations aimed at reducing and controlling emissions of GHGs[1]. The special case of water vapour is handled in a separate paragraph.

The first systematic observation of GHGs was started in 1957 when Ralph Keeling, a professor at the Scripps Institution of Oceanography, began measurements of amount fractions of atmospheric CO_2 from Mauna Loa, Hawaii and Antarctica [13]. Keeling's measurements were based on a locally maintained calibration scale that he could show was highly stable over decades. Since, the measuring system has expanded considerably to a partnership of sites operated by 100 countries, covering an increasing range of GHGs. The responsibility for providing a stable scale has also now been taken up by the WMO GAW CCLs. Observation communities and metrologists have increasingly collaborated, through the WMO-CIPM collaboration agreement formalised in 2010, and a joint Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology - the Working Group on Gas Analysis (CCQM-GAWG) group organises biannual meetings and regular intercomparisons of the measurement scales used by the WMO.

As well as the importance of measuring the amount fractions of GHGs to obtain reliable observations for climate modelling, anthropogenic GHG gas emissions are also collated on local and national scales for the emission inventories that are required to be reported annually to the UN Framework Convention on Climate Change. Reporting of national GHG emissions currently takes the form of bottom-up inventories generated from activity data and emissions factors for each species, both of which have uncertainties. Atmospheric measurements provide a means of independently verifying these estimates, an exercise routinely carried out only by the UK, Switzerland and Australia [14] and recognised as good practise by the Intergovernmental Panel on Climate Change (IPCC) [15]. This verification can also guide improvements to inventory compilation and reduction in uncertainty and requires reliable data from a network of measurement sites, which is combined in an atmospheric transport model to relate the measured amount fraction to an emission flux. Centralised calibration for each species is essential to minimise biases between measurement sites and accurate assessment of

---

[1] For example, fluorinated halocarbons are regulated in the regulation (EU) No 517/2014 (F-gas regulation); the Kyoto Protocol, developed under the United Nations Framework Convention on Climate Change (UNFCCC); the Montreal protocol, Kigali amendment, rules for emission inventories developed under the UNFCCC, EC directives, ICOS and the WMO GAW programme.

observation uncertainties is necessary for emissions estimates. Infrastructure and reporting are already in place for many species, but work needs to be done to reduce and better quantify measurement uncertainty and to improve traceability between networks.

GHG measurements are typically performed for long durations by automated instruments – often at remote locations in the case of background studies. For this reason, the instrument itself is not regularly calibrated by an NMI/DI, as is stated in the survey participants' responses. The instruments are instead calibrated in the field following standard community-accepted procedures often with community-accepted reference materials. These procedures are peer-reviewed (3 responses) and confirmed through intercomparison (4 responses) but are usually not subject to formal accreditation; for example, only one respondent laboratory indicated that they were accredited to ISO/IEC 17025. However, there remain many situations where different networks use differing routes of traceability (SI or community based) and this can make comparisons and consequently interoperability between networks and observations more challenging.

3.1.1.1 CO2, N_2O, CH_4 and Isotopes

As countries increase their commitment to emissions reduction, there is a growing need for GHG emission source and sink attribution in addition to the total atmospheric burden; for example, discriminating anthropogenic from natural sources by measuring ratios of stable isotopes and of tracer species. Such measurements have previously been deployed for targeted campaigns but are only now being installed at atmospheric monitoring stations for long-term measurements as field-deployable instruments are developed, with new requirements for traceable reference materials and calibration strategies. Consequently, there is an urgent need to provide a validated measurement infrastructure and provide the basis for stable, metrologically traceable and comparable measurements. Certified reference materials for these components are required with challengingly low amount fractions and associated uncertainties to improve the QA and control in the global networks such as those coordinated by the WMO. For instance, the network compatibility goals (= maximal persistent bias between measurement records [16]) are for CO_2 of 0.1 µmol/mol in the northern hemisphere and 0.05 µmol/mol in the southern hemisphere at typical ambient amount fractions (380 µmol/mol to 450 µmol/mol) and 0.01 ‰ for $\delta^{13}C$-CO_2 and 0.05 ‰ for $\delta^{18}O$-CO_2 for isotopic ratios.

The WMO has therefore stated the need for new traceable isotopic reference materials for atmospheric measurements and set compatibility targets [16]. The *European Metrology Programme for Innovation and Research* (EMPIR) project 19ENV05 STELLAR, started in 2020, will address these needs. In particular, the project will provide improved CO_2 references meeting the community needs and will develop the first ever CH_4 gas reference materials linked to the VPDB scale, though further work is expected to be required to meet the target uncertainty defined by the DQO from WMO-GAW. Validation routines and traceability chains for spectroscopic techniques will also be developed to allow state-of-the-art measurements to be made in the field.

Case Study: Nitrous Oxide (N_2O) Isotopes

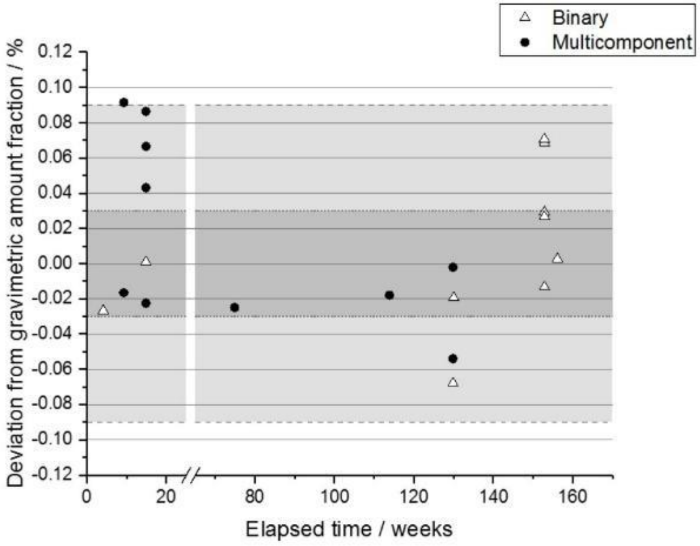


Figure 3: Difference between gravimetrically-prepared N_2O mixtures (with amount fraction in the range 320 nmol mol^{-1} to 360 nmol mol^{-1}) and measurement against a reference standard over 3 years. Open triangles represent mixtures in synthetic air and filled circles represent mixtures with an atmospheric amount fraction of CO_2, CH_4 and CO. The dark grey shading represents the WMO-GAW compatibility goal and the lighter grey shading indicates the extended compatibility goal.

Nitrous oxide is the third most important long-lived GHG after carbon dioxide and methane. N_2O has a relatively small inter-hemispheric gradient and a strong annual cycle due to its connection with agriculture. The temporal and spatial gradients require high measurement precision and comparability between measurement sites to understand the sources and sinks better. New optical instruments have the potential to improve the measurement precision and analyse the isotopic composition, which will provide new insights into sources, sinks and help distinguish natural and anthropogenic sources. New reference materials were needed with an isotopic composition that also achieves the WMO-GAW DQOs. Under the EMPIR-funded Metrology for stable isotope reference standards (SIRS 16ENV06), new reference mixtures have been produced for N_2O for a range of isotopic compositions at ambient amount fraction and have demonstrated stability of the mixtures over three years within the WMO-GAW expanded compatibility goal of 0.3 nmol mol^{-1} and the $\delta^{15}N$ and $\delta^{18}O$ value of the reference materials was also demonstrated to be stable with reducing pressure.

3.1.1.2 Radon

Radon is a naturally-occurring radioactive species with a short half-life emitted almost homogeneously from land that can be used as a tracer for the origin of an air mass reaching a monitoring site [17]. Radon data can be used to improve uncertainty evaluations in atmospheric transport models and emissions estimates through the radon tracer method (RTM), which uses correlation between GHG and radon concentrations. Climatic Atmospheric Monitoring Networks (AMNs), like the European Integrated Carbon Observation System (ICOS), are infrastructures that operate GHG monitoring stations and include atmospheric radon monitors in their stations. However, this radon data needs significant improvement in terms of the accuracy of radon activity concentrations in the range found in the atmosphere: between 1 Bq m^{-3} and 100 Bq m^{-3}, and traceability for the network of observations.

At present there is no traceability chain for atmospheric low-level radon activity concentration measurements and there is a need for improved transfer standards to bring traceability to the in-situ instruments. Furthermore, the application of radon as an atmospheric tracer of transport and mixing is dependent on accurate knowledge of its exhalation rate from the ground, which is dependent on the distribution of the precursor isotope, radium, in the soil and time-dependent properties such as porosity and soil moisture. This requires the development of maps of surface radon flux activity.

The EURAMET project EMPIR 19ENV01 traceRadon is responding to these needs by establishing a traceability chain for atmospheric low-level radon activity concentrations and a transfer standard to provide calibration to a network of measurement sites. This project is only a first step in establishing SI traceability for atmospheric low-level radon activity concentrations and validating the RTM. Further work will be needed to implement this method in environmental networks like ICOS and to link to other data sources (EU JRC geological maps and online radiological measurements). Reference data sets will increasingly play a role in the validation process in order to enable the use of artificial intelligence in a quality assured manner. Traceability for radon activity concentrations below 1 Bq m^{-3} is a challenge for the ICOS measuring stations at or on the water and developing suitable methods for this would be the next metrological step in the expansion of RTM.

The long-term goal is to use RTMs to gain access to information on natural and anthropogenic GHG emissions. Validation of methods like RTM are of special importance, because GHG concentration measurements in the atmosphere do not provide information about GHG fluxes between the compartments (here from soil to atmosphere), but there is an access to such systemic observations by means of radioactive tracers. A similar approach could be envisioned for the atmosphere-water cycle and the atmosphere-soil-vegetation cycle. Here iodine isotopes and Tc-99 could be used if radionuclide metrology can be developed in this environmental domain with sufficient uncertainties and small decision threshold and detection limits. This will also give the opportunity to perform a comparison of RTMs with other GHG attribution studies to improve understanding of both methods.

3.1.2 Halogenated Compounds

A growing contribution to radiative forcing comes from synthetic GHGs – gases with few natural sources. These were identified in a 2012 publication [18] as responsible for a growth in radiative forcing 19 % as large as that from CO_2 since the pre-industrial era. A major category of synthetic GHGs is halogenated compounds, with uses such as refrigeration, foam blowing and as blanketing gases in metal production. Many of these species are also ozone-depleting and are being phased out under the Montreal Protocol. The atmospheric burden of the replacements, however, is growing and these new compounds have high radiative efficiencies and long lifetimes [19]. Trends in these gases have been measured by the Advanced Atmospheric Gases Experiment (AGAGE) since 1978 at remote “background” locations around the world [20]. These measurements are referenced to a suite of primary standards maintained by the Scripps Institute for Oceanography; however, the new compounds are present at very low amounts and have no current standards, which makes such measurements and traceability provision difficult. Furthermore, for halogenated compounds, no CCL has been appointed by WMO GAW and for some newly emitted compounds no standards exist at all [21]. The European Metrology Research Programme (EMRP) project ENV52 HIGHGAS already achieved good progress on improving measurement standards for some of these compounds (in particular for carbon dioxide, CH_4, N_2O, and halogenated compounds such as SF_6). However, a lot of work remains, such as developing new traceable reference materials for halogenated compounds lacking reference and fulfilling criteria to host the CCL for these compounds.

3.1.3 Aerosol and Ozone Precursors

Ozone and aerosols contribute to the radiative forcing and therefore directly affect the climate. They are monitored along with the precursor species that react in the atmosphere to form ozone and aerosols. One main type of precursors is the broad family of volatile organic compounds

(VOCs), such as oxygenated compounds and terpenes. These compounds are regulated by several international directives and/or treaties, e.g. [22], [23]. To control the effectiveness of these treaties and to assess climate and air quality trends, the amount-of-substance fractions of these substances need to be monitored. The availability of high-quality stable reference materials and well-defined methods is essential to reach the necessary accuracy, as expressed in the WMO DQOs, as well as other monitoring networks; e.g. to achieve ACTRIS measurement uncertainties less than 10 % (k=2), the calibration uncertainty needs to be less than 5 % [24], which, as can be seen below in Section 3.1.3.1, is far from being achieved.

This topic was also raised during the survey and the webinar organised within this EMN, with one respondent writing:

> “Current calibration methods are sufficient for isolated field deployments. For PTR-MS [Proton Transfer Reaction – Mass Spectrometry] long term measurements and inter-comparability of different instruments, more standardized protocols and gas standards need to be developed and applied. Sampling artefacts due to surface effects in the inlet are not well constrained. This is a problem for many oxygenated molecules.”

Another scientist working on these gases requests support from metrology institutes:

> “Knowledge transfer with respect to trace gas - material interaction, i.e. ultimately with respect to stability of reference gases.”

3.1.3.1 Volatile Organic Compounds

For stable VOCs, such as Non-Methane Hydrocarbons (NMHCs), stable calibration gas mixtures are available as well as best practice guidelines, (e.g. the measurement guidelines by ACTRIS/GAW). However, this is not the case for many oxygenated VOCs and terpenes due to their reactivity and their low amount of substance fractions, which are translated into storage, surface effect and measurements artefact issues. Furthermore, current VOC reference gas mixtures are at a higher concentration than in the atmosphere and in a different matrix (e.g. nitrogen instead of air). Therefore, there is a need to develop fit-for-purpose reference gas mixtures, close to ambient air level and matrix to ensure an unbroken SI-traceable calibration chain.

An EMPIR Project (19ENV06 MetClimVOC) started in June 2020 to continue previous work and address the needs for selected “priority” VOCs.

Case Study: Volatile Organic Compounds

The abundance of atmospheric VOCs is low (parts-per-trillion (pmol/mol) to parts-per-billion (nmol/mol)) with some components being very reactive, which creates challenges for producing the reference materials, sampling ambient air and analysing them.

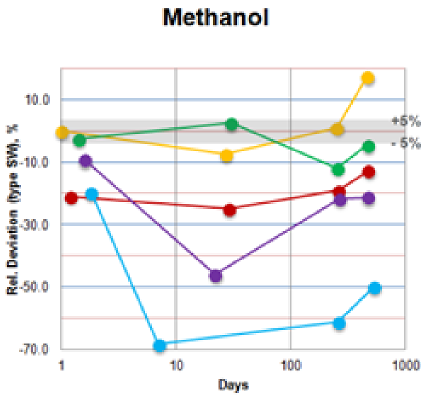


Figure 4: Stability of methanol gas mixtures in different cylinders types at nmol/mol level over about 2 years. The relative deviation represents the difference with the filled concentration at day 0. The grey area indicates the DQO of the GAW for methanol.

Substantial progress has been made during recent years. For example, new traceable reference gases were established [25], mobile, dynamic reference gas generators were developed, VOC-free zero gases were obtained [26] and new coatings for tubing and fittings that minimise adsorption and desorption effects are available on the market have been tested (e.g. EMRP JRPs ENV56 KEY-VOCS). However, some of the DQOs are currently not met for all specified compounds or they are still being assessed. Moreover, the WMO GAW implementation plan 2016 – 2023 [12] states as a key activity that "uncertainty calculation" and "full traceability to the primary standard" for all measurements reported is needed. Currently, no WMO GAW guidelines exist for the mentioned classes of VOCs.

3.1.3.2 NO_2

Nitrogen dioxide (NO_2) is a reactive and toxic gas, which plays a key role in ozone and secondary particle formation and influences the oxidative capacity of the atmosphere. As a consequence of NO_2 impacts on air quality and climate change, governments have been urged to develop effective mitigation strategies. To support these efforts, more than 3000 sites in Europe are reporting NO_2 data. However, the most common analytical method used is chemiluminescence (CLD) − an indirect measuring method − which compromises the accuracy of measurements. Recent advances in selective NO_2 measurement techniques highlight the need for characterising and evaluating their applicability for long-term NO_2 measurements, as well as for generating appropriate reference gas standards of NO_2 at atmospherically relevant concentrations with low uncertainties (≤ 1 %) and long stability (≥ 2 years), that comply with the WMO GAW DQOs defined in their implementation plan 2016 – 2023 [12].

Since June 2017, major progress towards addressing these needs has been made within the framework of the EURAMET EMPIR project Metrology for nitrogen dioxide (16ENV05 MetNO2; [27]). The project aimed to develop traceable static reference standards for NO_2 (1 µmol/mol – 10 µmol/mol) with a target uncertainty ≤ 0.5 % ($k = 2$) and stability ≥ 2 years and high accuracy traceable dynamic reference standards (10 nmol/mol – 500 nmol/mol) with at target uncertainty ≤ 1 %. Another project objective was to validate selective spectroscopic methods for direct NO_2 measurements. For dynamic reference, the target of 1 % was achieved for primary generation in the lab under well-controlled conditions. Work remains to disseminate these reference gas mixtures in an easy way in the field, since the uncertainty achieved during MetNO2 in the field was between 2 % and 3 %.

As there is substantial added value to be derived from atmospheric NO_2 measurement data that has been primarily collected for regulatory compliance, there is a need to communicate the benefits of making direct measurements of NO_2 to stakeholders as opposed to those inferred from the indirect CLD method. It is also necessary to provide them with the means to carry out these measurements through the provision of a standard reference method that is not based on the CLD technique and the provision of stable NO_2 reference materials at ambient relevant amount fractions. A new work item on a standard reference method for direct NO_2 measurements is being developed in CEN/TC 264/WG 12 and requires the support of the metrology community to provide insight on the state of the art for uncertainty assessment and traceability.

3.1.4 Aerosol Properties

3.1.4.1 Black Carbon Mass Concentration

Carbonaceous particles have recently received increased attention from the scientific community and policy makers. Black Carbon (BC), in particular, is of high importance because of its direct role in climate change and as a measure of combustion-generated aerosol for air quality purposes. The lack of a metrological framework for particulate BC absorption measurements was highlighted in 2013 by the European Environment Agency (EEA) as well as by the CCQM-GAWG Particulate Workshop [28] in April 2015, and this has led to the development of the CCQM-GAWG roadmap and EMPIR project 16ENV02 Black Carbon.

The 16ENV02 Black Carbon project, in close collaboration with the stakeholders, has brought a clearer metrological focus to measurements of aerosol absorption coefficient, measured in Mm^{-1}, at one wavelength, which gives the metric known as Equivalent Black Carbon. Further work is needed to extend this to the mass concentration of actual black carbon, measured in $\mu g/m^3$, to multiple-wavelength instruments, and to refine and improve the calibration procedures in place. More specifically, there is a need:

1. To measure, using SI-traceable techniques, the BC mass concentration and optical absorption properties of different types of aerosols found in ambient air. This will improve capabilities to identify and control sources of pollution, and also allow measurements to be reported as BC in $\mu g/m^3$ with realistic conversion from the aerosol absorption in Mm^{-1};
2. To extend the scope of work on traceability from single-wavelength to multiple-wavelength instruments. Calibration factors are expected to differ at different wavelengths for a complex mixture of reasons. Full traceability will lead to much improved wavelength-dependence analysis of aerosol particle properties, greatly improving their benefits for diagnosis, such as distinguishing between BC and brown carbon;
3. To refine the calibration procedures for field ambient instruments, to improve accuracy, portability and decrease cost;

3.1.4.2 EU Regulation

Currently, particulate matter (PM) mass concentration is the only regulated aerosol metric [29]. PM mass concentration, whilst useful, is not the most informative metric to characterise the potential of particles to cause climate change or the disparate detrimental health effects reported in the literature. The focus on mass also precludes the application of intelligent targeting of "climate-relevant" and 'health-relevant' constituents. There is a need to define new metrics, in particular beyond PM2.5 and PM10 (diameter smaller than 2.5 and 10 µm), which consider the complex chemical composition and optical properties of PM.

3.1.4.3 Reference Aerosols

In order to calibrate and validate in situ aerosol measurements, there is a need for reference aerosols that simulate the physiochemical properties of real-world aerosol mixtures.

Ambient aerosols are typically a complex mixture of inorganic and organic constituents and, in Europe, are typically present as "fresh" or "aged" (i.e. organically coated) soot from fossil-fuel burning, inorganic salts such as ammonium sulfate and ammonium nitrate, mineral dust particles and water. Different aerosol mixtures are also needed to represent more general aerosols, for example, the combustion particles from biomass burning, or general source-specific reference aerosols, such as those from ship or aircraft emissions.

Carbonaceous particles from ship emissions are typically complex mixtures of elemental carbon, organic matter, sulfuric acid and metal particles. Aircraft emissions are dominated by small combustion particles with diameters of the order of 30 nm; much smaller than those emitted by diesel engines (e.g. ≈ 90 nm). In addition, freshly emitted combustion particles from aircraft engines exhibit high elemental carbon mass fractions (> 80 %). It is currently highly challenging to produce such reference aerosols in the laboratory with commercial combustion generators.

Within the 16ENV07 Aeromet Project [30], a new facility was designed for the generation of ambient-like aerosols in the laboratory under controlled conditions. But currently no reference aerosols exist for the broader source specific aerosol mixtures. Once fit-for-purpose aerosol mixtures are available, one of the next challenges will be to define and calibrate transfer standards to be used further in the field.

3.2 Water Vapour

Water vapour is the main GHG as it is responsible for about 60 % of the natural greenhouse effect. Water vapour condenses to produce clouds, changing the atmospheric radiative properties and releasing latent heat that drives the atmospheric circulation. There are two ECVs for water vapour – an upper air water vapour ECV that relates to water vapour in the tropopause and stratosphere, and a surface water vapour ECV, that relates to humidity near the surface, which affects evaporation and the strength of the hydrological and energy cycles.

There are vigorous ongoing discussions within the research community on whether stratospheric humidity has changed in response to anthropogenic warming and how any further change is expected to influence the Earth's energy budget. At the same time, water vapour measurements, particularly around the tropopause, are known to have large measurement uncertainties. Even key mechanisms governing humidity in this region are not fully understood, leading in turn to significant deficiencies in the predictive power of global climate models. Currently, satellites and research-quality instruments on aircraft and balloon platforms are the main sources of humidity measurements around the tropopause, and differences between these measurement systems have been difficult to reconcile.

Measurements in the upper troposphere/lower stratosphere (UTLS, circa 5 km to 30 km) are a challenge because of the extreme ambient conditions. Nevertheless, the measurement of water and temperature are of particular importance for climate trend studies and have to be routinely executed. In the upper troposphere and lower stratosphere, the temperature could be as low as

-80 °C and the pressure reaches 1000 Pa. Water vapour is present with typical fractional amounts ranging from 1 µmol/mol to 10 µmol/mol. Comparisons of different commercial and research-based hygrometers show significant disagreement, whether or not they are based on similar or different working principles, with disagreements up to 100 % for very low levels of water vapour (below 2 µmol/mol) [31]–[33]. These discrepancies are far larger than the current ECV required measurement uncertainty of 5 % (k = 2) for profiles of upper air water vapour. In addition, the guide WMO-nº 8 [34] specifies the requirements for specific humidity in the UTLS for the application area of climate, indicating a goal uncertainty of 4 % (k = 2).

Water vapour can be measured in terms of dew-point or frost-point temperature, as relative humidity, as an atmospheric composition measurement (amount of substance), or via satellite microwave limb sounders. Requirements for satellite limb sounding are discussed as part of the general remote sensing discussion in Section 6.

3.2.1 Water Vapour Through Amount of Substance

To reach the measurement uncertainty requirements, improved reference and measurement techniques are needed. For instance, laser-based techniques are being considered as a viable alternative to the Cryogenic Frost-point Hygrometer (CFH), which is currently considered to be the reference method. There is an important and urgent practical step to be taken in instrument design, as the cooling agent (CHF_3) required for CFH operation is to be phased out by legal agreements and there is, therefore, a major, worldwide challenge to ensure the continuation of the high-quality observation of this key ECV [31], [32]. To validate and calibrate such an instrument, here the accuracy (or even the availability) of the spectral line parameters and the applied fitting model, SI-traceable reference gas mixtures with a very low amount of substance and a well-defined uncertainty are needed. Because of the properties of water (e.g. adsorption, reaction), the production process of the reference gas mixture is particularly challenging Several projects are working on improving the measurement of low water amount fraction. A new project started in September 2020 funded by GCOS-CH (Swiss funding of GCOS).

3.2.2 Water Vapour Through Measurement of Relative Humidity

Relative humidity (RH) is the humidity variable most commonly observed at near-surface level. The RH of a humid gas is usually defined as the ratio of some humidity quantity to the same quantity at saturation at the same temperature. This is the definition accepted by the WMO since 1950 ([34], Chapter 4, Annex 4.a, pp 1.4–27); however there a number of similar definitions in widespread use with resulting ambiguity, and work is needed to establish a fundamental basis to support one definition over another. An equally serious and related problem is the inability of the WMO definition (and of most alternative definitions) to cover the full range over which other humidity quantities apply and relative-humidity sensors respond usefully [35]. Therefore, the definition of RH needs to be revisited as it should have a firm thermodynamic foundation, be unambiguous and consistent. In addition, the definition should be linked to a practical realisation allowing measurements traceable to the SI. There is not even an internationally agreed symbol for RH. In addition, reliable observations using electronic relative humidity sensors can be difficult due to effects of condensation, contamination, ageing and temperature effects, and relative humidity itself is highly sensitive to temperature. Observations of RH are routinely converted to express other humidity quantities, including dew-point and frost-point temperature, water partial pressure, and amount or mass fraction or ratio, which are used in numerical models, in comparisons with other measurement types, and for their explanatory power in climatological and meteorological textbooks and research articles.

Atmospheric water vapour observation is based on both in situ and remote-sensing measurement techniques with different measurement uncertainties, collocation and representativeness errors. For applications in synoptic, aeronautical, agricultural and marine meteorology, hydrology and climatology, the WMO- nº 8 Guide recommends target uncertainties for surface humidity of 1 %rh and 5 %rh (both $k = 2$) at high-range and mid-range RH, respectively, with a reporting resolution of 1 %rh ([34] Table 4.1, p I.4-2 and Annex 1.B), whilst recognising that these can be difficult to achieve in practice. This can only be achieved with an improvement of the primary standards and the traceability chain together with a metrological characterisation of the humidity sensors, including a thorough investigation of the sources of uncertainty in the measurements.

Best-performing state-of-the-art radiosondes with application of sophisticated correction schemes were demonstrated to achieve a mean relative uncertainty of about 1 % in the lower and middle troposphere (where RH is more than10 %rh) and of about 2 % in the upper troposphere under mostly night-time conditions [33]. For instance, in some cases a repeatability of 0.01 %rh and a standard uncertainty of 1 %rh has been reported for RH determinations down to -70 °C for the standardised frequencies method on a research radiosonde. However, in view of the multitude of different radiosonde types and correction schemes in use, the required accuracy for climate research is still very difficult to realise in operational radiosonde services. Facilities for calibration at low temperature and pressure are needed to develop a procedure mimicking as closely as possible the environmental conditions encountered during the ascent flight and are being developed by the observational communities. Metrology institutes are supporting the development of such facilities.

3.3 Near-Surface Atmosphere Measurements

The WMO encourages its members (National Meteorological and Hydrological Services, NMHS), to use adequate measurement procedures to provide robust and comparable observations of near-surface atmosphere ECVs, traceable to the SI. According to WMO [36]

> “The primary quality factor of a measurement is the set of intrinsic parameters of the instrument used.”

Focusing on air temperature, a consistent measurement uncertainty calculation needs a complete knowledge of the measuring system, starting with the behaviour of this set of intrinsic parameters (response time, self-heating, etc), in addition to other external influences such as the place where the measurements are performed and the influence of other meteorological parameters like precipitation and solar radiation (from Annex IV of [37]):

> “Environmental conditions of a site may generate measurement errors exceeding the tolerances envisaged for instruments […]. It is often environmental conditions that distort results, influencing their representativeness.”

A correct measurement of air temperature is still a measurement challenge due to the complexity of thermodynamic processes involved in the heat transfer from the air and the sensors. Quantifying the effects of the many quantities of influence, such as air speed, radiation, condensation and convection is not easy. It is also not easy to develop an uncertainty budget for air temperature measurements. Moreover, thermometers are calibrated in liquid, where adiabatic conditions are very different from those met in the field (air), making the calibration procedure less representative of the in-field measurement conditions. Finally, even the measurand is not formally defined, in terms of a practical definition of air temperature (should this be about dry air, in steady state at zero radiation or at what extent of such conditions).

The CIPM Consultative Committee for Thermometry (CCT) strategy document [38] describes the need to improve the quality of calibration of thermometers in air. Previous activities in this direction have already been undertaken within EMRP projects MeteoMet and MeteoMet2, and EMPIR project 17SIP02 SimpleMeteoU.

3.3.1 Surface-Based Measurements in Extreme Environments, Key Climate Regions and Challenging Conditions

Some regions of the world, particularly the Arctic and Antarctica, as well as high mountain regions and coastal regions, are particularly sensitive to climate change and are at risk from sudden, irreversible changes. The recent IPCC Special Report on the Oceans and Cryosphere [9] highlighted the risks to these regions and their importance to humanity. In the technical summary it says:

> “Long-term sustained observations and continued modelling are critical for detecting, understanding and predicting ocean and cryosphere change, providing the knowledge to inform risk assessments and adaptation planning (high confidence). Knowledge gaps exist in scientific knowledge for important regions, parameters and processes of ocean and cryosphere change, […]”.

Comparable, reliable and accurate environmental measurements in these key regions are therefore needed to enable the early detection of climate trends. Higher accuracy measurements in many remote locations are needed to capture these trends. However, on-site observations in these extreme environments are limited by the logistical challenges of the remote location, the extreme conditions the sensors are exposed to and the difficulty in recalibrating the instruments or understanding their performance under these conditions. Environmental conditions, such as those in high mountains, can have a large impact on instruments in terms of introducing errors and increasing drifts. As an example, the backward albedo radiation from snow-covered surface can introduce errors of the order of 1 °C in temperature readings, as evaluated during the MeteoMet2 project [39]. Further challenges are in providing ongoing SI-traceability to instruments in these locations. The MeteoMet project also prototyped transfer standards that could bring traceability to remote locations such as high mountains (see also Section 5.3 on permafrost measurements).

In its “Call to Action”, the WMO High Mountain Summit in 2019 [40] describes the need to establish an “integrated high-mountain observation, prediction and services initiative with user-centred goals”. They urge governments to;

> “...address critical gaps in mountain Earth system observations in order to support integrated predictions and services, giving priority to the strengthening of remote-sensing observations of the mountain cryosphere and to the development of intra- and inter-operability of data platforms of operational and research programmes and projects, upon which services are built,”.

Developing interoperability of data from a broad range of observational techniques requires robust metrological uncertainty analysis of observations (instruments, environmental conditions and data processing) as well as comparisons. In the 16th session of the WMO CIMO [41], there was a recognition of the importance of comparisons in these difficult locations. Comparing observations from different techniques requires uncertainty analysis on both techniques, as well as a determination of the mismatch in the comparison – the extent to which the two observations measure different things (e.g. the remote sensing methods will average over a large area, while local conditions may vary rapidly in mountainous terrain). As in other fields, a collaboration between observation experts and metrologists, and a collaboration across technical disciplines is necessary to understand and resolve the specific challenges of these delicate environments.

High mountains are characterised by delicate environmental equilibria. A multidisciplinary approach is therefore required to understand better the amplified effects of climate trends in such areas. Metrology of physical parameters should interact with biological and geological observations, extending the interest also to flora and fauna observations. Dedicated procedures are required to establish comparability and cross feed of data for studies and activities in the mountains. An interdisciplinary and multidisciplinary approach is therefore required, where metrology for ECVs of atmosphere and cryosphere will contribute to the overall understanding of the evolution of the alpine environment.

Similar challenges are met in urban areas. The determination of urban climate includes a lot of difficulties, since the environment surrounding the instrumentation changes rapidly in both space and time. Measurements are influenced by these variations in a complicated way, generating a complex associated uncertainty budget, which is not robustly understood and making comparisons of different instruments and observational methods difficult.

Metrologists can support the establishment of traceability, uncertainty analysis and interpretability of observations in both remote and urban areas, and it is important that metrologists engage with and participate in the expert multidisciplinary communities that work to understand these environments.

3.3.2 Surface Reference Networks and Data Records

Climate observations have, in general, come from meteorological stations. But such observations were not originally taken for climate purposes and systematic step-changes in records from individual meteorological stations, as measurement technologies and methodologies evolve (e.g. introduction of compact screens, transition to automated measurements), represent one of the main issues linked to obtaining climate record information from long meteorological data series. When aggregated, these changes result in spurious step-changes and trends in global and regional meteorological records, which must be corrected to reconstruct a faithful climate record accurately. The data management of existing meteorological observations is also fragmented with the absence of a coordinated global programme for data rescue and provision, data management, data curation and data usage [42]. As a consequence, there are many emerging needs in agriculture, transport or energy which are ill-served. In addition, there is an increasing need for high-quality discovery and observations metadata: indicators of quality and uncertainties, in addition to known changes in measurement techniques, practices, locations and siting. These data and metadata are essential in provision of scientifically robust climate services [43].

In the future, it is reasonable to expect that observing networks will continue to evolve as technology improves and user requirements increase. It is possible that such changes will prove difficult to homogenise and would thus threaten the continuity of existing data series. Anticipating such future changes, the GCOS and the WMO requested a group of scientists to develop an outline for a global land surface climate FRM network. [44]. The WMO INFCOM approved the resolution 4.1.1(4)/1 (INFCOM-1) for the development of a draft implementation plan for the GCOS Surface Reference Network (GSRN). Reference quality observations are directly traceable to the International System of Units (SI) standards and include full documentation of all components of their uncertainty. Such observations respond to the need for monitoring the changes that occur in the climate and ensure greater confidence in the assessment of future climate change and variability. Such a network will also support timely political decisions around mitigation and adaptation. A GSRN will contribute to the improvement of:

- The current climate observing system, by enabling well-characterised time series that can be used with confidence at network sites
- Instrument performance that transfers down to other broader global regional and national networks;
- The calibration/validation of satellite data;
- The understanding and validation of models.

In parallel, the WMO is currently defining a Global Basic Observation Network (GBON) to meet the requirements of Global Network Weather Prediction, including re-analysis in support of climate monitoring. This network has a pragmatic approach and the connection to the future “reference” network would be welcomed (see documents at [45]). There are clear requirements for metrologists to continue to engage with the communities establishing these (reference) networks (including those with expertise in atmospheric composition and land ECVs, especially vegetation ECVs; see also Section 5 and Section 6.3). Sustained collaboration with the metrological community is key to the development and maintenance of a long-term reference quality network.

3.3.3 Measurement Comparability

The WMO Instruments and Methods of Observation Programme (IMOP) establishes in its terms of reference the duty of organising comparisons [34]. The role of metrology in such comparisons was recognised by the WMO Commission for Instruments and Methods of Observation (WMO-CIMO)[2] in its 17th session (October 2018) [46], where the importance of traceability and comparability of measurements was highlighted and WMO-CIMO pushed forward needed actions for the improvement of both. The metrology community has already started working on these needs (EMPIR project 19SIP06 COAT) in air temperature measurements.

Case Study: Weather Data for Climate Change Studies

The WMO is responsible for ensuring and harmonising measurements across global networks, so all results are comparable and traceable to the same standard and have links to the SI. To do so, it needed to develop methods to test the ability of calibration laboratories to perform reliable and consistent calibrations. Interlaboratory comparison (ILC) serves as a tool for comparison of measurement results carried out by accredited or non-accredited calibration laboratories in the relevant field of measurement. ILC represents a very effective means to demonstrate technical competence of participants and also serves as a technical base for accreditation. Furthermore, it is the most important element for monitoring the quality of measurement results as required by ISO/IEC 17025:2005 standard for laboratories in part 5.9.

The EMRP Project Metrology for Essential Climate Variables (MeteoMet-2) promoted, prepared and organised a large ILC for calibration laboratories of 19 European National Meteorological and Hydrological services. The work was performed in collaboration with

---

[2] WMO is in the process of restructuring and CIMO, the Committee for Instruments and Methods of Operation, is one of the committees of the former structure that is still operational in a transition period

the WMO Regional Instrument Center (ARSO-Slovenia) and the MeteoMet part. The ILC results have been published as a WMO IOM report and showed the validity and importance of having linked calibration procedures and standards among all the participants. The success of the activity was such that at the end of the MeteoMet project the WMO proposed to extend the ILC to other regions. Using the same equipment and procedure a linked ILC was organised in Regions II and V (Asia) with the National Meteorological Agency of Japan (JMA) as pilot. The ILC has now been extended to Africa and South America, with the vision to cover all the world, for a global comparability of laboratories and observations in atmospheric measurements for near surface parameters temperature, pressure and humidity.

As a result, climate scientists will have the confidence they need to use near surface meteorological data to inform ongoing climate monitoring and trend predictions. Feeding this additional data into climate models will improve their accuracy and help governments make better informed decisions about the best ways to mitigate and adapt to climate change.

3.4 Upper-Air Observations of Atmospheric Properties

The international community has recognised the need for high quality upper-air climate observations and the GCOS has established a reference measurement network with the aim of providing CDRs throughout the atmospheric column into the stratosphere. The GRUAN was implemented in 2008 (see [47]) building on existing operational and research facilities utilising state of the art ground-based instrumentation. The task of GRUAN is to provide long-term, highly accurate measurements of the atmospheric profile with emphasis on the upper troposphere and lower stratosphere (upper-air). The specific demand is to define and assure reference quality for the data and data products, prioritising the GCOS ECVs: temperature, wind speed and direction, water vapour, cloud properties, and Earth radiation budget (including solar irradiance). A formal definition of reference quality for GRUAN measurements has been established (see [48]) including the full traceability of all measurements to SI units or internationally accepted standards, uncertainty analyses, which distinguish contributions from systematic and random error, comprehensive documentation, data validation, and metadata collection and management. With the quality level of the GRUAN data, results from more spatially-comprehensive global observing systems including satellites and current radiosonde networks can be constrained and calibrated.

The scientific challenges identified by the GRUAN community [49] include:

- Characterisation of changes in ECVs, in particular temperature, humidity, and wind
- Understanding the climatology and variability of humidity, particularly in the region around the tropopause since this is where changes have their largest effect on climate sensitivity
- Understanding changes in the hydrological cycle
- Understanding and monitoring tropopause characteristics
- Understanding the vertical profile of temperature trends
- Bringing closure to the Earth’s radiation budget and balance
- Understanding climate processes and improving climate models

Particular emphasis is put on the challenge of temperature and water vapour measurements. It is recognised that existing records of upper-air temperatures are insufficient to meet the growing range of needs for studying climate. They generally lack continuity, homogeneity and representativeness of data because past measurements were seldom intended for long-term climate research, but rather for short-term weather forecasting. Instrumentation and methods for separating climate change signals from the inevitable non-climatic effects, caused by measurement biases, instrument instabilities and network inhomogeneities, are therefore essential.

The measurement issues for upper atmospheric measurements were also addressed within the Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring (GAIA-CLIM) European Horizon2020 project [50]. The focus of this project was to establish sound methodologies for surface-based and sub-orbital measurements (non-satellite measurements) to be used to help characterise and validate satellite-based EO data. Robust EO instrument characterisation needs to go beyond the simple coincidence criteria for a given set of EO and ground-based / sub-orbital measurements. It also requires quantified uncertainty estimation for the reference measurements and an understanding of additional uncertainties that result from the spatial and temporal mismatch between the measurements.

A key part of the GAIA-CLIM project was a Gaps Assessment and Impacts Document (GAID) which, through careful analysis against both existing and envisaged user requirements, aimed to identify unfulfilled user needs (‘gaps’) in the observation capability of upper air ECVs. This document is publicly available [51], and contains details of specific gaps. However, it is worth noting that the following topics were identified as high-level recommendations for future activity:

- Improve the metrological characterisation of a suite of non-satellite measurement techniques: Striving for traceable, reference quality, fiducial measurement series
- Instigate and sustain time-bounded access to a comprehensive set of harmonised fiducial reference data and metadata holdings under a common data model and open data policy that enables interoperability for applications
- Improve knowledge of fundamental spectroscopy and undertake associated innovations in radiative-transfer modelling
- Improve quantification of the effects of surface properties to reduce uncertainties in satellite data assimilation and satellite to non-satellite data comparisons
- Develop and provide tools that convert non-satellite fiducial reference quality measurements to Top-Of-Atmosphere radiance equivalents with associated rigorously quantified uncertainties
- Improve the basis for assigning co-locations and quantifying rigorously the associated uncertainties, including steps towards operational provision of colocation uncertainties

3.5 Summary of Needs for Atmosphere

Metrology Challenges for Observations of Atmosphere ECVs
Fit-for-purpose working standards at appropriate concentrations, ensuring an unbroken SI-traceable calibration chain. There is also a need to improve the sampling and analytical methods for ambient measurements and to assess relevant influence parameters (typically aerosol and ozone precursors such as VOCs or NO_2)**.**
Certified reference materials for newly emitted halogenated compounds lacking standards (e.g. greenhouse gases such as HCFCs). Reference materials and a traceability chain for new measurements of isotopic composition and atmospheric tracers (e.g. dissemination of N_2O) must also be established**.**
Development of calibration procedures for aerosol properties using filter-based absorption photometers, to provide a metrological framework for aerosol metrics beyond PM2.5 and PM10, and to generate source-specific reference aerosols in the laboratory.
Improve reference methods and instrumentation, typically for humidity measurements in the upper troposphere/lower stratosphere (less than 10 µmol/mol) under adapted environmental conditions (e.g. low pressure and temperature).
Support the establishment of the surface reference network by the GCOS (similar to GRUAN) as the top level of the WMO Integrated Global Observing System prescribed tiered approach in conjunction with the launch of the Global Basic Observing Network.
Metrological support for comparisons, particularly for challenging measurements lacking well-defined SI-traceability, and to compare surface, upper-air and satellite measurements (i.e. where there are very different traceability chains), including on site comparisons with metrological rigour for extreme environments and challenging locations e.g. cryosphere and high mountains.
Metrology support for specific initiatives focussed on the cryosphere, high mountains and urban areas; Need for metrologists to participate in multidisciplinary partnerships focussed on observations and predictions in key climate areas (e.g to participate in the establishment of an integrated high-mountain observation, prediction and services initiative).
Metrological support for field calibrations and measurements, including guidelines for using measuring devices on site including environmental influences and their uncertainty contribution.
Improved metrological characterisation of spectral parameters for chemical compounds (e.g. absorption cross-section, spectral line)**.**

4 OCEAN OBSERVATIONS

Today, ocean observation is internationally coordinated under the auspices of the GOOS, which is a programme executed by the IOC of the UNESCO. It is a sustained collaborative system of ocean observations, encompassing in situ networks, satellite systems, governments, UN agencies and individual scientists. The EOVs are identified by the respective expert panels of GOOS. Beyond this overarching structure, the marine communities collaborate in many national, European and international organisations, usually covering several EOVs. Generally, the observational infrastructures are well coordinated, essentially because they are multidisciplinary, and they must provide data serviceable to a variety of users and to policymakers. However, in contrast to comparable infrastructures related to atmosphere observation, which have a long heritage driven by the requirement to provide reliable data for weather forecasting, ocean observation (with a few exceptions, e.g. water temperature and sea height) has not had a comparable societal need in the past. Internationally-agreed concepts for QA of observation data have only rarely been established in ocean observation. The need for reliable ocean data for a range of additional parameters, i.e. in conjunction with climate change modelling, has caused a rethink within the last few decades.

Recently, several initiatives have been initiated to establish internationally-agreed QA criteria, best practice guides and standards [52]–[54]. An international group of experts endorsed by the International Oceanographic Data and Information Exchange (IODE) of the IOC states in their review paper [55] that:

> [52]–[54]“A first and foremost requirement for collaboration in ocean observing is the need to follow well-defined methods,”

and,

> “Best practices and standards are the two most common dimensions present in broadly accepted methodologies and serve to ensure consistency in achieving a superior product or end state.”,

> “Standards … may become mandatory legislated standards, such as the European INSPIRE[3] legislation.”

One answer to these needs has been the creation of the Ocean Best Practices System (OBPS) [54] which;

> “provides a foundation upon which the ocean community can more systematically develop and use best practices”.

Nevertheless, it was also recognised during the discussions at the first EMN stakeholder webinar that some countries have missing or badly distributed QA capacities, including data management capacities. Training is needed for countries with inadequate calibration capabilities and it will be important to improve coordination on data quality issues with these EU and international infrastructures, and to provide mechanisms to better utilise capabilities across borders.

---

[3] INSPIRE is the Infrastructure for Spatial Information for Europe, see https://inspire.ec.europa.eu/.

Despite efforts to establish internationally-agreed quality standards, it still seems that fundamental metrological concepts are often not considered. For instance, quality criteria have been defined for EOVs by the relevant panels and the coordinated institutions of GOOS [56]. However, for some of these, instead of using traceability and measurement uncertainty to qualify measurement results, there are general requirements for the assessment of the measurement method in terms of its technical maturity and a flagging system to exclude bad data. Measurement uncertainty is often understood as the accuracy of a device as stated by the manufacturer without any evidence of traceability. For instance, the EOOS Strategy 2018 – 2022 document [57], states that:

> “[...] data do not always meet user needs: despite the availability of relevant European ocean data, many are not used for environmental assessments (e.g. with the context of the Marine Strategy Framework Directive) due to lack of data provenance, low quality control and accreditation.”

and that,

> “Ocean observation data collection is often neither standardized nor quality controlled to an agreed level.”

Moreover, suitable reference materials and common methodologies for uncertainty evaluations, including all influencing factors, are missing according to several respondents to our survey:

> “Metrology institutes could help scientific communities to understand the breadth of error sources in their data and how these can be quantified in uncertainty estimates.”

The EOOS Strategy and Implementation Plans [57], [58] further describe the importance of involving metrologists in developing “best practice” for the “systematic harmonisation of ocean observing in Europe”. The EU’s Joint Programming Initiative: “Healthy and Productive Seas and Oceans” (JPI Oceans) wrote in its report on the need for EOOS [59], that:

> “… to be useful for research and decision-making at a transnational level, all the incoming data have to be comparable and amenable to fitness-for-purpose assessments in relation to specific user-group requirements. This will require measurements to be metrologically referenced […]”.

At present, such metrological referencing is limited to some EOVs. Referring to biogeochemical variables, the JPI Oceans report states:

> “although a few recognized standards are in place, no certified reference material is available … to this end, the National Metrology Institutes can contribute by helping the oceanographers and manufacturers to establish validated metrological procedures”.

In fact, there is, apart from the EMN for Climate and Ocean observation, poor coordination on metrological issues for the in situ oceanographic sector in or outside Europe which addresses the need to establish metrological concepts. Regional monitoring communities are instead organised around Regional Sea Conventions (where available) and regional operational oceanography is organised around ROOS (under EuroGOOS). On the other hand, some respondents of the survey stressed the benefits of the interaction between the metrology community and the oceanographic community, but stronger interactions were requested:

> “I have interacted with the metrology community, and the interaction has been useful, stimulating, and constructive. In my view, metrological institutes need to improve their interaction with the communities involved in measuring activity outside their specific sphere of competencies, building the links and collaborations needed to address practical problems relating to metrology and measurement.”

and,

> “Some better transfer of knowledge to the user from the metrology community would be useful to be able to ’speak the same language’ and to identify the most important uncertainty sources. An easy way to compute uncertainty would be welcomed. I would welcome further engagement with the metrology community.”

To ensure data quality with respect to internationally accepted quality standards, a few oceanographic institutes are implementing metrological principles in their QA systems. In some cases, they are (considering) embedding the QA of EOV measurements into existing national accreditation systems that, in turn, are embedded in European and international accreditation structures like the European Accreditation (EA) and the International Laboratory Accreditation Cooperation (ILAC). Institutions seeking accreditation for the measurement of specific EOVs would have to give evidence that their measurements are performed in compliance with international (written) standards, e.g. ISO/IEC 17025 for calibration laboratories or ISO 22013 for sensor producers, which include metrological requirements like traceability and adequate uncertainty calculations explicitly. Besides some typical physical measurement quantities also measured in other areas, accreditation bodies are rarely prepared to cover EOVs that are based on measurement procedures specific to oceanography. As a consequence, oceanographic institutes, accreditation bodies and respective standardisation bodies seek metrological support on those issues.

It must be emphasised that the landscape of ocean observation is rather diverse, all the more so as there are currently over 30 EOVs. While metrological concepts are well established in some oceanic institutions, others are not even aware of them. Therefore, the general needs of ocean observation roughly generalised above cannot readily be applied to any EOV. In fact, each ocean measurand must be assessed individually in this regard.

4.1 Physical Variables in the Ocean

Variables such as absolute salinity, temperature, ocean currents, are widely measured and have significance in the context of some of the global concerns of the moment: climate change and circulation.

The metrological challenges for measurements of some key climatological observables (salinity, pH and relative humidity) have been summarised in a series of four highly recognised papers [33], [35], [60], [61] and presented below. Besides other technical issues, the determination of appropriate uncertainties, often mentioned together with unresolved traceability issues, has been identified as one of the most pressing metrological needs in oceanography. This need has also been expressed by other initiatives related to oceanographic observation practice.

4.1.1 Absolute Salinity

Absolute salinity is a term used to quantify the total mass of substances, i.e. salts, dissolved in pure water to form a given mass of seawater. In the form of latent heat, the oceans export 50 % to 90 % of the absorbed solar energy to the atmosphere by evaporating water. The related global hydrological cycle is reflected in the distribution of sea-surface salinities; arid regions in the trade-wind belts show higher salinities, and humid regions at the equator and at mid-latitudes have lower salinities than the global average. Salinity deviations affect the density gradients in the ocean and in this way modify the worldwide marine ‘conveyor belt’ of heat transports. Along with temperature and pressure as key parameters for ocean modelling and observation, salinity significantly influences almost every property of seawater, including its heat capacity, sound speed, refractive index and viscosity. Thus, local long-term trends in salinity are important indicators for climatic changes in the terrestrial water cycles.

Direct Absolute Salinity measurement methods, i.e. measuring the mass of dissolved salts, are inappropriate for the frequent regular measurements required in ocean observation, since they are labour intensive, i.e. they are difficult to automate for in situ use and they have relatively large uncertainties. In practice, oceanographers, for many years, have therefore used the fast, reliable and robust technique described by the Practical Salinity Scale of 1978 (PSS 78) to approximate salinity. This so-called “Practical Salinity” is defined by using proxy measurements of electrical conductivity relative to that of a bottled reference material called International Association for the Physical Sciences of the Oceans (IAPSO) Standard Seawater (SSW) However, even though Practical Salinity is widely measured, there are still some metrological challenges in using it as a measure for Absolute Salinity, the quantity that is required for oceanographic and climatological models.

Thus, to establish SI traceability for Absolute Salinity in practice, it is necessary to provide SI-traceability to Practical Salinity measurements, i.e. to the Practical Salinity of the reference SSW. A recent European metrological research project “ENV05”, established SI traceability for SSW, thereby solving the former problem of Practical Salinity to guarantee long-term comparability of the measured values. However, while Practical Salinity can be measured with sufficiently small uncertainties, the associated uncertainty increases significantly if the results are used as estimates for Absolute Salinity. This problem becomes even more critical if seawater with composition anomalies is investigated (e.g. from marginal seas or coastal waters). Thus, there is a fundamental need to investigate and reduce the uncertainty of obtaining Absolute Salinity values from Practical Salinity measurements.

Practical Salinity is measured over a wide temperature and pressure range relevant in oceanographic practice. However, calibration of Practical Salinity sensors is basically done with respect to a single set of temperature, pressure and salinity values of SSW, that is 14.996 °C, normal atmospheric pressure and Practical Salinity of 35 (dimensionless quantity). Up to now, the assumed reproducibility of Practical Salinity results relies on three measurement series performed under laboratory conditions during the establishment of PSS-78. There have been many efforts to establish best practices for sensor calibration and qualitative flagging systems to qualify measurement methods of Practical Salinity at varying temperatures and pressures. However, the actual, quantified uncertainty of in-field Practical Salinity measurements, considering the complete traceability chain, is unknown to date. There is a need to quantify the uncertainty of in situ Practical Salinity measurements under varying temperature and pressure conditions.

Finally, a huge number of salinity measurements, and likewise temperature and pressure data, are measured in situ over extended periods of time by Conductivity, Temperature and Depth (pressure) probes (CTD) in various ways, using Argo floats, ships, buoys, platforms and other systems. The large numbers of such instruments, the near impossibility of recalibrating many of them once deployed, and the wide range of environmental conditions that they are subjected to, mean that standard, laboratory-based uncertainty evaluation methods cannot be readily applied. Thus, there is a need for appropriate uncertainty calculation concepts that considers the specific conditions of ocean observation and ideally improved instrumentation/methods that can maintain or evaluate uncertainty of measurement following deployment.

4.1.2 Ocean Temperature

The temperature of the oceans is a major driver of the weather and climate. Sea surface temperature is a dominant component of the interchange of energy, momentum and gases between the oceans and the atmosphere, and as such forms a core input to numerical weather prediction models that underpin weather forecasting, monitoring for extreme weather events (hurricanes and cyclones), climate modelling and oceanography models. It is also an essential input for other applications: a core component for ecosystem assessment of fish abundance and for sensitive biodiverse environments such as coral reefs, as well as being important for tourism, fisheries, disaster monitoring, transport and environmental policy. Subsurface temperature measurements are needed to understand the uptake of heat by the ocean (the oceans are estimated to have absorbed 90 % of the anthropogenic heating of the Earth), and therefore to understand global ocean circulation, stratification and coastal shelf exchange processes.

Operational data and long-term CDRs of sea surface temperature are predominantly provided by satellite sensors in both low Earth orbit and geostationary orbit; the metrological needs for such sensors are described in Section 6. In situ observations from drifters, moored buoys, floats and ships are used for validation (and retrieval[4]) of satellite observations and through historical observations can extend the record of sea surface temperature back to the 1850s. In situ observations are also essential in providing information about the temperature profile in the upper layers of the ocean. While satellites only measure the top few microns/millimetres of the sea surface, the “skin”, in situ sensors can measure “subskin”, “near surface”, “foundation” (insensitive to diurnal temperature variations) and “deep ocean” temperatures.

In situ skin sea surface temperature is measured with voluntary observing ships (also known as “ships of opportunity”) carrying infrared (IR) radiometers. Other in-situ temperature measurements are based on contact thermometers, usually as part of a CTD package that is also used to measure practical salinity (Section 4.1.1). Such systems are mounted on drifters, moored buoys, Argo floats and on voluntary observing ships.

In general, the uncertainty associated with a contact thermometer’s measurement in a liquid is of the order of millikelvin, and such instruments can be well-calibrated prior to deployment. One exception to this is deep sea sensors which operate under high-pressure (see case study on deep sea sensors). For near-surface and bulk temperature measurements, the dominant sources of uncertainty arise from our knowledge (or lack thereof) of where the sensor is, whether it may have changed since deployment and how representative that temperature measurement is of the desired ECV/EOV. In the upper layers of the ocean (particularly in the top few metres) the temperature varies rapidly with depth and is sensitive to solar radiation. Models are used to correct observations at different depths (remote sensing of the top few microns, thermometers a few centimetres or metres below the surface) to a common reference depth. Such models have inherent uncertainties and rely on auxiliary information (e.g. sea surface state, wind speed, solar radiation levels) that is itself uncertain. Furthermore, the exact depth of the instrument may no

---

[4] “Retrieval” is the process used to convert the satellite-measured top-of-atmosphere radiance in a few spectral bands to obtain a sea surface temperature, accounting for the radiance of the atmosphere. Some retrieval algorithms use in situ observations to provide a prior state, whereas others are based entirely on satellite observations and use in situ observations for independent comparison.

longer be known, for example, surface-floating drifters initially are held at a constant depth by a drogue, but at some point the drogue will break off and the depth will change. For deeper sensors, the depth is established via the pressure measurement of the CTD probe, but the longitude and latitude of an observation may be difficult to estimate and relies on interpolating surface measurements using knowledge of ocean currents. Furthermore, instruments may get covered in barnacles or other marine life, and this will also affect the relationship between the measured temperature and the sea temperature.

Sea temperature communities have considerable experience of handling these different issues and use metrological methods and data science techniques to identify outliers, step-changes in instruments (e.g. drogue loss on a drifting thermometer), and to compare in situ and satellite observations. They are also already collaborating strongly with the metrology community. A recent workshop organised as part of the Fiducial Reference Measurement (FRM) for Surface Temperatures from Satellites (FRM4STS) project [62] recommended:

> “Research on the means to improve traceability and trust in measurements from floating buoys whilst maintaining relatively low costs.”

The workshop also demonstrated the interest in methods for recovering drifting buoys after deployment to improve the evaluation of instrument changes over time. Many of these requirements were driven by the need for FRMs for satellite observations (see Section 6.3).

Case Study: Deep-Sea Sensors

The target standard uncertainty set by the World Ocean Circulation Experiment (WOCE) World Hydrographic Program (WHP) on ocean temperature measurements is 2 mK. Such an uncertainty level is necessary to ensure traceability of long-term sea temperature measurements, since recent studies have reported an increase of 5 mK per decade for the temperature of deep ocean water in the North Pacific Ocean.

However, deep-ocean in-situ temperature measurements may suffer larger uncertainties, especially because of the effect of water pressure on thermometers, which can introduce deviations of several millikelvin at pressures up to 60 MPa. An investigation has revealed the pressure effect, but measurements realised under extremely well-controlled temperature and pressure conditions are necessary.

In addition, considering that the most widespread deep-ocean thermometers are based on thermistors, there is a need for metrological validation of the temperature-resistance linearisation equation adopted – especially on high-grade deep-sea reference thermometers – in order to assess their ability for providing uncertainties at the millikelvin level or below. The only laboratories able to reach such uncertainty levels are NMIs. The JRP ENV58 MeteoMet2 has realised a comparator block to be used for pressure dependence investigation of deep ocean thermistors carried out for temperatures in the range 0 °C to 10 °C and pressures in the range 0.1 MPa to 60 MPa.

4.2 Biogeochemical Variables of the Ocean

4.2.1 Inorganic Carbon

One of the key issues with respect to climate change is the ocean acidification (OA) phenomenon.

> "Ocean acidification is an emerging global problem. Over the last decade, there has been much focus in the ocean science community on studying the potential impacts of ocean acidification. Since sustained efforts to monitor ocean acidification worldwide are only beginning, it is currently impossible to predict exactly how ocean acidification impacts will cascade throughout the marine food chain and affect the overall structure of marine ecosystems [60],[63]."

The ocean absorbs about 30 % of the CO_2 that is released in the atmosphere, and as levels of atmospheric CO_2 increase, so do the levels in the ocean.

When CO_2 is absorbed, a series of chemical reactions occur resulting in higher seawater acidity and causes carbonate ions to be relatively less abundant. Decreases in carbonate ions can make building and maintaining shells and other calcium carbonate structures difficult for calcifying organisms because these ions are important building blocks of structures such as sea shells and coral skeletons.

The EOV monitoring OA is Inorganic Carbon, which itself is described by four sub-variables: Dissolved Inorganic Carbon (DIC), Total Alkalinity (TA), pH and Partial pressure of carbon dioxide (pCO2). The measurement of at least two of the four sub-variables are needed to determine Inorganic Carbon.

4.2.1.1 pH

In 2016, according to Dickson et al. [60]:

> "quantitative understanding of exactly what we are measuring is significantly worse than the repeatability that can be achieved by particular measurement techniques. […], there is as yet no single recommended measurement procedure, nor is there an internationally accepted reference standard for seawater pH measurement that enables different laboratories to achieve comparable measurements reliably."

Moreover, for pH,

> "technical issues are particularly problematic in seawater studies. First, seawater has a high ionic strength, which causes problems when using conventional pH calibration standards. Second, some current research problems such as the detection of the long-term anthropogenically-driven changes in ocean carbon chemistry over multi-decadal timescales would benefit from an extremely small standard uncertainty in pH measurements (as small as 0.003), albeit over a fairly narrow range of pH, and this is far smaller than the differences between many of the available operationally-defined 'pH' quantities. The notation 'pH' in quotation marks is used here to emphasise that, although commonly called pH, these various operationally-defined quantities are not identical to the accepted definition."

Since the publication of that review, efforts have been made to harmonise "pH" measurements. The IOC-UNESCO has put in place an indicator (SDG 14.3.1) associated to the SDG 14.3 target: "Minimize and address the impacts of ocean acidification, including through enhanced scientific cooperation at all levels". SDG 14.3.1 – Average marine acidity (pH) measured at agreed suite of representative sampling stations – provides guidelines on how to ensure data quality.

Case Study: Ocean Acidification

Faced with the critical need for better defining the measurement requirements, the Global Ocean Acidification Observing Network (GOA-ON) distinguished a “weather goal” and a “climate goal”. These 2 goals are differentiated by the level of uncertainty requirements. For pH measurements, a level of standard uncertainty of ± 0.02 pH is associated with the “weather goal” and considered necessary to identify relative spatial patterns and short-term variations. A standard uncertainty level of ± 0.003 pH is needed to support detection of long term (multi-decadal timescales) trends enabling to address “climate goal” challenges. If the latter objective concerns mainly the open ocean water, the former can be of interest for marine coastal environments whose role in the global carbon budget is currently under debate.

Based on the developments of the JRP ENV05 “OCEAN”, a set of reference materials of Tris/Tris-HCl at pH_T values of 7.6 and 8.2 and expanded uncertainty of 0.005, which fits for the “climate goal” purpose, has been produced as proof of concept. This set of RMs has been used to organise an interlaboratory comparison in the frame of the JPI OCEANS to evaluate the comparability of different pH_T measurement devices.

However, it is recognised that despite harmonised methods and best practices, uncertainty estimation tools are still lacking. For example, traceable pH buffers in seawater and brackish waters, as well as purified pH dyes are not commercially available at present. According to our stakeholders needs survey, regular provision of pH reference materials through a European institution would be extremely helpful:

> “For pH measurements: Traceable pH buffers in seawater and brackish waters, as well as purified pH dyes are not commercially available”.

4.2.1.2 Partial Pressure of Carbon Dioxide – $p\mathrm{CO}_2$

CO_2 naturally exchanges between the atmosphere and the surface of the ocean. The CO_2 content of the upper ocean has been increasing in parallel with the CO_2 in the atmosphere. Some of the CO_2 that dissolves in seawater remains in the form of a dissolved gas that can freely exchange with the atmosphere and be taken up directly by marine plants and phytoplankton. This fraction is usually referred to as “dissolved or aqueous carbon dioxide”, and it is typically expressed as the partial pressure of CO_2, $p\mathrm{CO}_2$. In general, the values of the $p\mathrm{CO}_2$ in seawater are slightly lower than the mole fractions of CO_2 in the atmosphere, and their fraction depends mainly on the temperature and relative humidity. For both the environmental compartments, i.e. air and water, there is a pressing need to ensure metrological traceability and obtain comparable results on spatial and temporal scales.

At present, $p\mathrm{CO}_2$ is one of the few measurable variables of the seawater carbonate system for which autonomous, in situ sensors, amenable to networking and capable of relaying measurements in or near real-time, are commercially available. These kinds of sensors are based either on the equilibration of a carrier gas phase with a seawater sample and subsequent determination of the CO_2 that diffuses through by means of non-dispersive infrared (NDIR) spectrometry (e.g., PSI CO2-Pro, Contros Hydro-C), on reagent-based colorimetry (e.g. SAMI-CO2) or on species-specific solid-state detectors.

Within the Joint Action of the JPI Oceans “European Marine Sensors Calibration Network”, $p\mathrm{CO}_2$ was selected as one of the four marine parameters to be implemented in a permanent pan-European calibration grid, to support the activities of marine observatories [63].

Due to the nature and diversity of the spatial and temporal scales over which observations of pCO_2 needs to be maintained in order to be useful, the implementation of proper networks of sensors has become a predominant feature of environmental and climate-related monitoring activity. Such sensors are beginning to be used more and more in different settings and under diverse conditions in operating contexts that range from straightforward monitoring to pure research. There are, however, several major hurdles hampering their widespread use. One of our survey respondents wrote:

> “In oceanography sensors, we need more standards for sensor calibration […]. Sometimes sensor calibration with the manufacturer does not have the precision required for ocean climate measurements.“

For marine pCO_2 measurements, the proper calibration of sensors and the regular assessment of their performances during use, represent critical steps in determining the suitability of generated data for climate-related applications. At present, there is also a limited number of certified reference materials easily available, some of which are strongly dependent on the instability of the matrix and on pH variation. There is a need for more reference materials, in order to rely on enough references both for calibration and for QC checks, considering that whenever a reference material is used for calibration, it cannot be used for QC. Furthermore, there is no existing harmonised and standardised method for the evaluation of the measurement uncertainty of the calibration of marine pCO_2 sensors, and the metrological traceability of the measurements they provide is not currently ensured by NMIs or accredited calibration laboratories.

4.2.2 Particulate Matter

Sand is an important resource, used in construction and in maintaining coastlines across northern Europe. As sources of sand on land become scarce, more sand is being extracted from the North Sea. Offshore sand extraction by dredging is strictly regulated, taking place in restricted areas of sandbanks, and is closely monitored to gauge the impact of sand extraction on the marine environment. The direct impact on the depth of the seabed can be mapped using a Multi-beam Echo-sounder (MBES), and the effects on the seabed and the unsustainable nature of marine sand as a resource are now well understood.

The indirect impacts are, however, still a topic of research. Extracting sand using trailing suction hopper dredgers generates sediment plumes of suspended particulate matter (SPM). Re-sedimentation of these plumes may impact the regions surrounding dredging sites – which can include areas of high biodiversity – which has consequences for the marine ecosystem. It is important to evaluate and quantify such indirect impact of sand extraction to guarantee the sustainability of sand extraction sites and their surrounding environment.

While sediment plumes generated by dredging operations are well studied, there have been few attempts to quantify the volume of SPM produced. SPM plumes could potentially be visualised and quantified using MBES, but this requires in situ measurements of SPM concentration simultaneously. Innovative solutions should be proposed to measure in real time and on-site particles suspended in the water column. This raises several metrological challenges, such as sensor calibration, dynamic operation and uncertainties of the derived measurements, the main aim being the creation of a validated scattering model that allows sediment information to be inferred from measurement data.

4.3 Biological and Ecosystem Variables of the Ocean

The demand for data on sea and ocean conditions is growing, particularly for biological and biodiversity datasets. The climate community has until now treated biological ECVs mainly as part of the carbon cycle, however, there is a broader need to understand change of life in the sea, in particular how diversity and abundance of life affects climate parameters and vice versa. According to the responders to our survey, research is needed:

> "to fill the spatial gap analysis of where physical, biogeochemical and biological EOVs/ECVs are measured together."

and,

> "Improved increase in policy relevance of measured biological ocean ECVs, realised by increased ability to reliably provide global synoptic data. This will require at least a doubling of effort over the next decade, accompanied by improved quality control and fair and open data".

Four responders to the survey indicated they were performing work on plankton (phyto- and zoo-) and two on marine habitat properties.

The use of multiple data sources for climatic studies would need consistency among the different measurements, considering sensor uncertainty. A greater number of sensors of good quality are needed to meet the frequency and stability requirements.

On the likelihood of metrology institutes providing more support for ECV record production, QA or application, respondents stated their view that it is not yet clear how this could be possible:

> "I don't know because I don't know what metrology institutes can do regarding the challenge to implement observing systems more broadly, in areas where they are needed but where socio-economic needs are substantial (developing nations, areas where environmental observations are not of interest politically)."

and,

> "The issue is more the disconnect between different data producers and/or user communities, resulting in different styles of working and a very broad set of requirements."

4.4 Ocean Sound

Acoustic energy propagates so well in the ocean that it is often the most effective way to probe the marine environment and communicate over long distances. This makes sound a critical feature of the ocean environment for marine life as well as for seagoing humans. Ocean sound has long been part of ocean observation systems, and has (relatively recently) been recognised as an EOV [64],[65]]. In addition, acoustic technology continues to provide the primary imaging and communication modalities for the exploration and exploitation of the ocean, underpinning key sectors such as oceanography, offshore energy extraction (including both oil and gas and marine renewable energy), and security applications.

Case Study: Ocean Sound

Low Frequency Traceability

Low frequency acoustics and vibration phenomena in the ocean are used to detect major natural events such as earthquakes, tsunamis and volcanic activity, are used in the study of large ocean basin scale phenomena (acoustic tomography for ocean currents, and the study of storms) and in determining the effects of climate change on sea temperature (ocean acoustic thermometry) [66]. It is also used by the International Monitoring System (IMS) of sensors to check compliance with the Comprehensive Nuclear-Test-Ban Treaty. However, the low frequency ranges used for detection (below 25 Hz) are not well covered by current measurement standards, limiting the reliability of data obtained. These challenges are being addressed by a new EMPIR project called InfraAUV (19ENV03) [67].

Sonar Quantitative Imaging

Sonar imaging is used to remotely detect and identify seabed objects, determine seafloor properties, and quantify benthos. However, there is a need to make an absolute comparison of the performance of different sonar systems for seabed imaging because of the use of different vessels, sonars, operators and settings. It is difficult to compare images and determine true changes because the imaging process typically involves normalisation processes which are often nonlinear (resulting in the loss of absolute information) [68], [69]. To facilitate comparison, a number of methods can be used. A reference patch of seabed is sometimes used with the assumption that this “reference” seabed does not change. Another solution might be adoption of “quantitative imaging” where the system whole performance is calibrated, the idea being (essentially) to ascribe an absolute quantitative value to the pixels of a (sonar) image, with the values having some meaning in terms of a physical parameter. This is a current topic in medical imaging (including using ultrasound) and is analogous to the need to calibrate remote satellite sensors used in EO. The method requires some kind of “calibration” of the overall transfer response of the imaging system, and can be done with standard targets [70]. A theoretical and experimental underpinning to support sonar imaging based on absolute, traceable measurement standards may result in a paradigm shift, where such an imaging capability would significantly improve the detection, classification, and quantification of seafloor objects, properties, and processes. It would also enable change detection, based on the comparison of images over time and with different sonar systems and operators.

Human activities have increased the sound levels throughout much of the ocean and acoustic noise is now recognised as a pollutant by international regulation and directives. Recent research has begun to uncover ways in which anthropogenic noise affects marine life, with the results triggering regulatory requirements in the Exclusive Economic Zones of many nations to reduce or manage the impacts of noise, which in turn drives research on effects. Current research is forging major advances in our understanding of how ocean sound is evolving and how different anthropogenic sources affect marine life, and how acoustic monitoring can be used to assess biodiversity and ecosystem health. Understanding the effects of ocean noise as a stressor requires estimating how ocean sound has changed historically, mapping sound throughout the oceans on a global scale over decades, and predicting sound fields that result from changes in the use of the oceans. The largest anthropogenic changes in the acoustic environment of the ocean are associated with industrialised coastal regions and major shipping lanes [71].

Furthermore, the characteristics of the acoustic environment evolve with transformations of the physical environment due to climate change. More extreme weather events affect sound generated by wind and waves in the oceans, and reduced sea ice affects the propagation of sound and changes the sound generating mechanisms in the Polar Regions (e.g. increased ice calving raises the sound levels in the adjacent seas). Changes in physical and biogeochemical ocean parameters will influence the ecosystems and migration patterns of fish and marine mammals and these climate-related changes can be observed by passive acoustic monitoring. The amount of sound energy absorbed by the oceans is affected by pH, and the degree to which ocean acidification will reduce absorption and increase ambient noise is under investigation. Sound propagates faster in warmer oceans, and this property can be exploited to make unique observations of climate change over large ocean volumes, including sea temperature [64]. Sonar imaging is used to map the seabed for habitat mapping and seabed classification, but comparisons of results are limited by lack of absolute benchmarking or calibration [69].

If ocean observation of sound is to be able to determine the status and changes in the above parameters, it requires traceability to internationally-validated standards, stable well-characterised sensors, and validated methods of comparison of different acoustic technologies. This presents several challenges for metrology:

- Traceability for absolute calibration is currently relatively weak, especially at low acoustic frequencies. This is true for sensors (hydrophones) at frequencies below 250 Hz where the availability of traceable calibrations is not widespread within the NMI community; but is particularly true for infrasonic frequencies (below 25 Hz) which are used for seismic detection, and some large-scale ocean basin oceanographic applications. It is hoped that progress will be made in the new EMPIR project InfraAUV (19ENV03).
- Traceability for deep ocean acoustic measurements is weak, with only two NMIs providing a service over a range of simulated depths and temperatures, and even these are only at hydrostatic pressures equivalent to ocean depths of up to 700 m.
- There is a need to develop better in situ calibration methods which would enable hydrophone performance to be assessed in the field, without the need for very expensive extraction for laboratory calibration.
- In seabed imaging, there is often a need to make an absolute comparison of the performance of different sonar systems used to image the sea floor. The use of different vessels, sonars, operators and settings makes it difficult to compare images and determine true changes because the imaging process is typically a relative and nonlinear one (resulting in the loss of absolute information) [68], [69]. To facilitate comparison, a reference patch of seabed is sometimes used, but another solution might be adoption of “quantitative imaging” where the system whole performance is validated using a standard target [70]. This is analogous to the need to calibrate remote satellite sensors used in EO (see Section 6.3).

4.5 Summary of Needs of Ocean Observations for Metrology

Metrology Challenges for Observations of the Ocean
Definition of proper measurands and fit-for-purpose high order and working standards that ensure unbroken SI-traceable calibration chains. Currently, some of the ocean ECVs and EOVs are not defined in terms of SI units (e.g. pH, salinity). This makes it difficult to compare results obtained in different time and places, particularly when technology breaks occur.
Certified reference materials are essential tools to ensure the metrological traceability of results via the calibration of instruments, or to validate analytical measurement methods. Currently very few reference materials exist for some of the ocean ECVs and EOVs (e. g. inorganic carbon variables, pCO2, TA, pH) and most of them are not certified by NMIs/DIs.
Development of a metrologically based QA/QC framework and associated tools to facilitate field measurement reliability and consistent uncertainties. Currently, few oceanographic institutions are familiar with ISO 17025 accreditation. A scheme could be created on the example of QA4EO, establishing guidelines written in collaboration between the oceanography and metrology communities (see case study on QA4EO and its implementation in Section 6).
Organisation of interlaboratory comparisons for in situ measurements following metrological best practice to establish 'degrees of equivalence' and biases to enable international interoperability and harmonisation for long term comparability.
Fit-for-purpose uncertainties for in situ measurements, including training courses: GCOS requirements set stringent target uncertainties for many of the ECVs which are close to the level of primary standards. In contrast to this demand, assignment of uncertainties according to metrological concepts is not well established in oceanography.
Moving beyond best practice guidance documents and standard measurement procedures to international documentary standards, which can provide longer stability of measurement procedures over time.
On-board calibration for underwater instruments mounted on research vessels continuously measuring oceanographic parameters such as temperature, salinity, pressure, sound speed and bathymetry to ensure traceability and accuracy of measurements over instruments' lifetimes and to account for environmental conditions and for their operation in dynamic mode.

5 IN SITU LAND ECV OBSERVATIONS AND NETWORKS

The land domain, while much smaller than the ocean and atmosphere in terms of area and volume, is host to most of humanity's activities and provides habitat to humans and to a significant fraction of the world's biodiversity. The land has a major role in the Earth's climate, particularly through the photosynthesis of plants, and the activities of humans. Table 5.1 lists the GCOS land ECVs. These cover the carbon cycle (biomass, photosynthesis, soil carbon and leaf area index, FAPAR, land cover as well as anthropogenic GHGs), areas particularly sensitive to climate changes (glaciers, ice sheets), radiation balance (temperatures, heat flux, albedo) and the water cycle (natural and anthropogenic water exchanges). Those that are shaded have a significant satellite-independent in situ measurement aspect with the darker shading indicating that in situ data provide the sole or dominant measurement method.

Table 5.1 GCOS listing of Land ECVs. Those shaded have a significant satellite-independent in situ measurement aspect with the darker shading indicating that in situ is the sole or dominant measurement method

Above Ground Biomass	Glaciers	Latent and Sensible Heat Fluxes
Albedo	Groundwater	Leaf Area Index
Anthropogenic Greenhouse Gas (GHG) Fluxes	Ice Sheets and Ice Shelves	Permafrost
Anthropogenic Water Use	Lakes	River Discharge
Fire	Land Cover	Snow
Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)	Land Surface Temperature	Soil Carbon
		Soil Moisture

The non-shaded ECVs in Table 5.1 are primarily measured by satellite and/or other remote sensing observation methods and whilst there are, of course, related in situ observations and networks to complement and validate these observations, for the purposes of this report these are included in the scope of Section 6. It should be noted that Section 6 does not, in general, address specific ECV needs but rather the general observational/metrological principles of satellite/airborne-based systems and other terrestrial-based remote sensing methods.

Although for the shaded, primarily in situ land ECVs there are some ECV-specific metrology needs, most can be summarised as needing support for the following:

- Establishing metrologically robust traceability chains to an internationally agreed reference and the associated uncertainty budgets
- Developing community consensus sampling methods
- Combining results from a network to establish a representative regional/global mean from which a long-term trend can be established

An initial review of these ECVs and their GCOS requirements concludes that most of the actual measurands are not particularly metrologically challenging and often follow well-defined protocols of an international network under the auspices of a body such as WMO. For example, Snow Depth and the related parameter Snow Water Equivalent are generally based on measurements made to a fixed reference 'depth measuring' pole at defined sites. In a similar manner, lake depth/area/volume is usually observed from gauge stations. These types of measurement cannot be carried out in anything close to a globally representative manner, but rather focus on time series of measurements at fixed locations. with satellite observations to complement changes in overall surface area. The land ECVs related to water have ISO standardised methods for monitoring, indicating relative maturity and are, in some cases, part of regulatory national reporting frameworks. Metrology institutes provide (and should continue to provide) ongoing support to such measurements through the calibration of instruments (e.g. thermometers, water speed flow meters for river discharge, etc) through normal calibration services and, where appropriate, support for standardisation (e.g. in the update of ISO3455 on flow measurement in open channels, which underpins river discharge measurements).

Some ECVs, particularly ‘Latent Heat and Sensible Heat Fluxes’ and ‘Soil Carbon’, whilst being important potential indicators, are still in the phase of defining suitable measurands and associated quality metrics. For these, it is possible that the metrology community may be able to provide input to help the expert community at this relatively early stage. However, the nature of these ECVs may not fit well within the current expertise of most NMI/DIs.

At this time the disparate nature of much of the land in situ ECV communities and the apparent lack of need for urgent metrological intervention will keep this section relatively short as those with most urgent needs relate to satellite observations, or in situ observations to provide fiducial references for satellites. These are addressed in Section 6. That is not to say there are no needs, but at this time the NMI community as a whole and this study in particular has not identified many specific urgent needs. As the EMN and its members mature in expertise, expand their community interactions and make more visible the value and willingness of the NMI community to help, there is likely to be an increased dialogue with those making land in situ observations and this may identify other specific needs in the future.

We have, at this time, identified the following ECVs where there has been some dialogue between the NMI and ECV community and/or where there have been some specific requests for support.

5.1 Anthropogenic Greenhouse Gases

This ECV concerns anthropogenic emissions of GHGs from industry, agriculture and waste disposal, as well as the contribution to GHG emissions from anthropogenic change in land use, e.g. through deforestation, agriculture methods and field usage (crop types, arable etc). Assessment of this ECV contributes to national declarations of emissions as part of the Paris Agreement 2015. The in situ based assessment is generally called the ‘bottom up’ approach, and satellite derived estimates ‘top down’. From a metrology context, one of the key challenges is to ensure that these two different methods are consistent.

The challenges that face the satellite observations and the remote sensing-based methods in general are included in Section 7 and those related to GHG measurements by in situ methods are discussed in Section 3.1.1.

5.2 Soil Moisture

Soil moisture is important to measure, not only from a climate perspective: drought, flood prediction, GHG exchange etc, but also for agriculture. In-situ soil moisture measurements are organised through the International Soil Moisture Network (ISMN) [69] which is a joint organisation of the international coordination bodies e.g. GEWEX, GEO, GCOS and CEOS with the aim of establishing a global database from the collation of data from various networks and observation stations. This database is intended to be the source of consistent quality-assured data against which change can be monitored. The data is also used by services such as QA4SM [73], which is used to validate satellite observations.

Although the ISMN has gone a long way to establishing harmonised QA and measurement procedures, it has requested support of the metrology community to rigorously determine and present uncertainties and traceability back to SI standards under field conditions with the view to create FRM-quality data sets (see also Section 6.3). There are, for example, ambitions to hold international comparisons to validate measurements made by different research teams and support on how to ensure a representative comparison and unambiguous analysis of results would benefit from NMI support. At the soil moisture calibration workshop in Madrid in 2016, discussion included the definition of the soil moisture measurand and the properties of sampling representativeness, as well as the depth-positioning of the instrument. In addition, traceability of sensors at the uncertainty levels needed for climate also needs support.

5.3 Permafrost

Permafrost is ground that contains water and remains frozen for at least two years (but usually millennia), and is a sensitive indicator of climate change. Thawing permafrost also releases GHGs (carbon dioxide and methane) into the atmosphere as well as destabilising landscapes with landslides. Monitoring change in depth and associated temperatures is a clear indicator not only of the onset of climate change but also as an early warning system for potential mitigation/adaptation actions. As well as the permafrost content of polar areas, which represents the majority of its extension on the planet, mountain permafrost is a sensitive indicator of climate trends in these delicate environments. In Europe, permafrost in the Alps is monitored by numerous stations to quickly capture its evolution, deeply linked with glacier retreat and disappearance. Accurate understanding of permafrost evolution in the Alps is therefore crucial to predict the permanence of ice content, which in turn is directly linked to fresh water availability: the impact of glacier retreat and permafrost disappearance is of high relevance on mountain environment, vegetation, agriculture, farming, and tourism. For polar permafrost, surface temperature can be measured from space (and is included in Section 7), while its small-scale local presence in the Alps requires on site monitoring of temperature profiles.

The Global Terrestrial Network for Permafrost (GTN-P), under the International Permafrost Association (IPA), coordinates hundreds of observation sites through common procedures and methods. Permafrost is monitored through temperature sensors in boreholes from 10s to 100s of metres deep with a required expanded uncertainty (and long-term stability) of ~0.1 K (k = 2) either during field campaigns, or with automated continuous monitoring. Work is needed to correct for convection effects within the borehole and to understand the impact of the sensors being chained together on the uncertainty. The permafrost Active Layer Thickness (ALT), defined as the surface layer of ground that freezes in the winter (seasonally frozen ground) and thaws in summer, is a GCOS ECV. When ALT information is obtained through an interpolation of temperature sensors data, along a vertical profile in the soil, uncertainties better than 0.1 °C are required to better predict permafrost presence in high mountains. To achieve long-term stability, the instruments need to be periodically recalibrated under field conditions, and therefore transfer standards that can be taken to field sites are required. During the MeteoMet-2 project, in-field calibration of such thermometers with metrological methods was prototyped. Best practice and guidelines are under development by the Global Cryosphere Watch, where metrologists are contributing to defining measurement procedures and evaluating uncertainties. The work still needs experimental investigation and comparison of methods and techniques to better fit recommendations on best practices and to reduce the overall measurement uncertainties below 0.1 °C.

See also the discussion on observations for key climatic regions in Section 3.3.1.

5.4 Summary of Needs of In Situ Land ECV Observations for Metrology

At this stage of engagement, we can summarise the needs for the ‘in situ’ Land ECVs as:

Metrology Challenges for In Situ Observations of Land ECVs
Support to establish and assess traceability and associated uncertainty to community agreed (ideally SI) references for measuring systems under operational conditions.
Support to write documentary measurement procedures and best practice in a metrologically robust manner.
Mathematics to facilitate representativeness of observation from multiple samples/sites at both single locations and as part of a network.

6 REMOTE SENSING FOR CLIMATE AND OBSERVING THE EARTH AND ITS ENVIRONMENT: SATELLITE, AIRBORNE, etc.

For many measurands, particularly where scales are dimensionally and geographically large, some form of remote sensing technique is required. Remote sensing techniques can use any part of the electromagnetic spectrum and may be passive (e.g. measuring the reflectance of solar radiation or emittance of IR radiation), or active, where the observation occurs due to the emission of electromagnetic radiation by the sensing system (e.g. LiDAR, RADAR, etc.).

Remote sensing techniques span across all ECV domains: land, ocean and atmosphere; and because their requirements are metrologically similar across these domains, they are all considered in this section and not the domain specific sections. Many metrological needs are independent of the observing platform and so are generalised and treated here together. Others are specific to satellite platforms, which have additional challenges, for example due to the inability to bring the instrument back into a laboratory for recalibration. More than 50 % of ECVs can only be monitored adequately using satellites, while a further 25 % require at least some satellite observations. Therefore, given the criticality of satellites in the climate observing system (and that the instrumentation and sampling in the broad sense are similar for remote sensing from any platform), satellite needs and examples dominate our discussion below, but the needs of those sensors flown in air and used in surface-based observations/networks, in addition to those used for satellite validation, are also very much in scope and key aspects will be highlighted in Section 6.4.

Some airborne and ground-based remote sensing activity is performed to provide local-scale post-launch calibration/validation (cal/val) or scaling for satellite data and thus has common goals and challenges to satellite observation. Furthermore, some in situ measurements and networks are established or adapted specifically to provide cal/val to satellite observation and these, along with efforts to ensure their traceability (to SI or appropriate community defined references), is also included here. This suite of satellite-optimised ‘traceable’ in situ methods and references are now being called fiducial reference measurements, or FRMs, by the satellite EO community. ESA in particular has initiated several projects dedicated to both creating FRMs and to undertake comparisons to ensure the world’s cal/val teams can evidence consistency in a metrologically-robust manner [74]. Some of these FRM networks may become linked to or part of the surface reference network discussed above in Section 3.3.2.

In addition to enabling ‘fit-for-purpose’, cost-efficient SI-traceable calibration and characterisation for the pre-flight/deployment of sensors, it is equally important to support the post-launch cal/val efforts. These efforts include in situ measurement devices, and also the means to scale between those, usually point measurements to the larger areas that are observed by the remote sensor – accounting for the observational path differences. In the absence of fully SI-traceable satellite sensors, given the shock of launch and harsh environment of space, it is difficult to rely solely on pre-flight calibration even if there is some form of on-board self-checking system, and so post-launch FRMs are critical. They are particularly important for climate studies, not only because the uncertainties required for climate are very demanding, and usually necessitate a time series of sensors over decades to build up sufficient signal to detect change, but also because of the need to assess orbital and scene-dependent environmental effects on the measurement system.

Remote sensing data are used for a wide variety of applications. While the sensors measure relatively simple quantities – the radiance or reflectance of the Earth, or the returned waveform from the active pulse, the measurand of interest (ECV) may need other information, parameters and models to retrieve the necessary information in a useable manner, e.g. any surface parameter needs to remove the effects of the atmosphere between sensor and surface. We therefore consider separately the metrological needs relating to the instrument and its observation (which describes what the satellite community calls “Level 1 products”, and sometimes, particularly for active sensors, “Level 2”) and the metrological needs relating to the evaluation of ECVs from those simple observations (what the satellite community calls “higher level products” and typically refers to “Levels 3 and 4” and, occasionally, Level 2).

Remotely sensed data (and indeed in situ data) are typically combined with model-derived information to provide evidence on the past, present and future state of the Earth’s climate and its ‘cycles’ (e.g. carbon, hydrological, radiation etc). This is necessary to understand the dynamics of climate conditions and their impact on societal systems. Climate data derived from satellites is often collected, processed and distributed by dedicated research centres and data warehouses that generate and distribute datasets with different spatial and temporal coverages for use by other researchers as well as public and private organisations. Given the critical importance of satellites to climate one key task of GCOS is to identify the principal observables, specific parameters (ECVs).

The detection of a trend from any of these ECVs requires many decades to build sufficient signal to detect above both instrumental noise, and uncertainty and variability due to natural variations in the environment itself. Thus, long-time base records of these ECVs are required, CDRs, and inevitably these are the result of combined data sets from multiple satellites over a long-time scale (Fundamental Data Records, FDR, or fundamental climate data records, FCDR, where it is considered adequate to meet a climate objective [75]).

It is important to note that for most ECVs, and indeed for other observed parameters/applications, the measurand observed by the sensor is not enough by itself. It often requires a combination of other data and/or models to retrieve the specific parameter of interest, some applying corrections for observing characteristics, e.g. spatial scale, illumination conditions, transmittance through the atmosphere others to link the physical measurand with bio-geophysical parameters e.g. amount of carbon stored in a forest or ocean, sea level height. Thus, in considering remote sensing it is necessary to be holistic and assess not only the sensor but also how the measured data is used and transformed.

6.1 Standardisation and Community Organisation

Satellite-based EO can be considered to be carried out by three groups of organisations:

1. The operational meteorological services are organised through the WMO and its GOS, and the Global Space-based InterCalibration System (GSICS) acts as the Cal/Val coordination group for space missions. Within Europe, the main organisation providing meteorological satellite data is EUMETSAT.
2. The public sector national and international space agencies come together through the CEOS, the space arm of the GEO, although membership of CEOS also includes many of the operational meteorological agencies. In Europe, the most active space agencies in CEOS are ESA, and the French (CNES), German (DLR), British (UKSA) and Belgian (Belspo) space agencies. In some countries there can be more than one civilian space commissioning body and the European Commission funds the Copernicus programme which includes not only satellite and in situ observations but also services and data centres for example the C3S. Many of the activities and associated needs of CEOS members can be found on its Cal/Val Portal [76].
3. More recently, there has been a rapid growth in commercial satellite operators who typically produce small sized satellites that provide very-high-resolution images for commercial applications. In Europe, the main organisations developing such satellites include SSTL,

Airbus, VITO, Thales Alenia Space, Clydespace and Planet amongst many others. There are trade associations for the commercial sector (satellite builders and data providers), both nationally and Europe-wide (EARSC.org) and these can provide a focal point for communications and common requirements. Currently within these trade organisations there is an increasing interest in some form of certification or QA programme for EO data and derived information as commercial markets develop.

There is increasing overlap between the three groups, with space agencies now providing more operational services (e.g. Sentinel missions and the operational Copernicus services), and with meteorological satellite data going back to the 1980s (and occasionally 1970s) being reanalysed to monitor climate trends. The reanalysis of older satellite data, and to a certain extent also current sensor data, presents a challenge because the pre-flight and onboard calibration processes were, and are, not in general designed for climate applications. Biases between sensors in a series need to be evaluated and corrected in a robust and consistent manner. Commercial satellite data are also being used for societal applications and their data purchased by government agencies. This blurring of boundaries between the data sets and move towards combining data from a broad range of satellites and their sensors is driven, in part, by the space agencies making their data open and freely available, a trend that started with the free release of USGS's Landsat data in 2008 (which led to economic benefit calculated as $2bn a year [77]). ESA made its EO data freely available in 2010. Today satellite data are also made available via Google Earth Engine and Amazon Web Services, opening access to an ever-growing number and range of users.

Case Study: QA4EO and its Implementation

The Quality Assurance Framework for Earth Observation (QA4EO) [78] was formally endorsed by CEOS in 2008 as a documented means to achieve the desired interoperability between satellite sensors and also broader non-satellite observations of the Earth as a whole. QA4EO is established around a guiding principle that "all data and derived products must have associated with them a quality indicator based on documented quantitative assessment of its traceability to a community-agreed reference standard, ideally SI units". The QA4EO guidelines that support this were written in collaboration between the EO and metrology communities, with NPL and NIST scientists developing the guidelines and helping translate these into concepts more suitable for EO communities. For example, the QA4EO guideline on comparisons (guideline 4) was based on the metrology community's MRA processes and, in particular, on the CCPR guidelines for comparisons [79]. The uptake of the concepts of QA4EO in the CEOS community has grown rapidly since 2008 as it has gained increasing recognition. ESA now explicitly requires the implementation of QA4EO principles in many of its processes.

There have also been many European research projects (funded by the space agencies ESA and EUMETSAT and by the European Union's Horizon 2020 programme and its predecessor) that have involved collaboration between metrologists and EO experts to develop methods for implementing QA4EO into each community. This has led to guidelines for developing and reviewing CDRs of ECVs [80], [81], to the establishment of FRM of SI-traceable ground truthing observations (e.g. [82]), and subsequent "FRM-networks", and to the development of guidelines for applying metrological techniques to

FCDRs (the basic satellite product that is used to generate different ECVs) [75]. It also underpins strategic vision papers such as that written by European Space Science Committee [83].

Despite the growth of applications which combine data from many sources, there is still a lot of work to do in order to establish a common reference frame that can be considered SI-traceable to fully enable an interoperable global Earth Observing system. GEO describes the implementation of a "Global Earth Observation System of Systems" (GEOSS) as its core mission. It is crucial for the implementation of GEOSS that data "are accessible, of identified quality and provenance and interoperable" [84]. In establishing the principles to achieve this, CEOS held several workshops and, in 2008, formally endorsed the QA4EO [78] as a documented means to achieve the desired interoperability (see case study box). The QA4EO principle and supporting guidelines were written collaboratively between the space agencies and metrology institutes and the principles have wide acceptance in the EO communities. The EO community is asking for wider support from metrologists in applying these methods to the full range of satellite sensors and their products. For example, in [80], which lists the "ten priority science gaps in assessing climate data quality", recommendation 5 says:

> "Since all satellite-derived ECV products start by using Level 1 data, the required measurand at this level should be derived by calibrating (on board and/or post-launch) the sensors or recalibrating them. In general, calibration errors usually present themselves in the form of biases in the Level 1 data when compared against trusted references. For satellite data, another challenge is that the pre-flight calibration may not be appropriate for in-orbit behaviour of the instrument. The metrology community should support the development of a framework for the metrological characterisation of satellite instruments that encompasses exploitation of ongoing pre-flight and post-launch calibration activities."

Some initiatives, such as the EU H2020 project FIDelity and Uncertainty in Climate data records from Earth Observation (FIDUCEO) [85], have made progress in this direction, developing methods not only to assess and remove biases but also to do so in a metrologically robust way through establishing and assessing the measurement equation for the sensor measured values, so that a full representative uncertainty budget for each data set (and time series) can be established and documented in a consistent manner. The FIDUCEO approach enables data sets from different sensors to be more readily merged into a long-time-base record as needed for climate applications.

It is a common theme in stakeholder community workshops, and also identified in our EMN stakeholder needs survey, for requests for the metrology community to engage on analyses for the full range of sensor types: solar reflective sensors (visible and shortwave IR), microwave sensors, thermal IR sensors, atmospheric sounding sensors (which use very high resolution spectrometry) and active sensors (radar altimetry, synthetic aperture radar (SAR), LiDAR). As one respondent to the survey indicated, even for modern sensors:

> "For active sensors no uncertainty for Level 1 is provided."

Metrology can provide a framework that encourages and facilitates the assessment and reporting of uncertainty for the full lifecycle of an EO data product: sensor pre-flight through to multiple-sensor-derived time series and collaborations between metrologists and sensor experts are needed to apply such a framework to all types of sensor.

In addition to the specialist needs of climate, the EO data providers are looking to expand the range of users of EO data and also to ensure that different data providers can have comparable products. Thus, the agency and commercial communities are working towards defining ways to establish and define the "fitness for purpose" of different data sets for different applications and to make the data sets as accessible as possible. In response, there has been an introduction of concepts such as "Analysis Ready Data" (ARD) and the provision of "data cubes", wherein different data sets are provided together for any particular location. There are currently active conversations in these communities about how to standardise such definitions, although in most cases these stop short of formal (e.g. ISO) standardisation. These communities have welcomed the involvement of metrologists as they implement more formal standards and bring a metrological robustness into both formal standards and guidelines and recommendations.

In encouraging standardisation we must also be careful. The CEOS working group for climate has not yet accepted the FIDUCEO project definition of a "fundamental climate data record" (FCDR) because of a fear that an overly strict emphasis on metrological completeness would mean that existing datasets would lose their status. Similarly, some ARD applications do not require radiometric calibration and associated uncertainties. In QA4EO, a similar balance led to the adoption of the phrase "quality indicator" rather than "uncertainty statement", although the term "quality indicator" was defined such that, wherever possible, an uncertainty statement was expected.

The current mechanisms for discovering and accessing ECVs present several challenges and barriers, and coordinated efforts from multiple disciplines and the integration of diverse data infrastructures and data are needed [86]. The needs of the different communities can be highly varied. However, the establishment of long-term and coherent time-series of global ECV products stored in CDRs should be seen as an opportunity and necessity to encourage dialogue between the GCOS, the remote sensing (EO) and the metrology community.

Nightingale *et al* [81] propose a framework for Evaluation and Quality Control (EQC) of climate data products derived from satellite and in situ observation to be catalogued within the C3S Climate Data Store (CDS). The CDS aims to make it easier for users to access complex climate datasets and turn them into useful information products. The EQC framework supports the C3S, implemented by ECMWF, as part of their operational QA programme with the purpose of presenting collected QA evidence in a standardised manner on each of the individual data sets.

Further action is required to ensure the quality of climate data sets derived from satellite and in situ observations and to provide users access to the range of information necessary to select relevant products for their specific applications confidently. Some of the key challenges [80] that the metrology community need to address can be summarised as follows:

- Support validation methods and uncertainty evaluation
- Establish and document the SI or (community agreed) traceability chain and provide meaningful quality flags to aid the user to fully interpret 'fitness for purpose' of the data in a consistent and readily accessible manner
- Representativeness, uncertainty and comparability of retrieval and correction algorithms and associated ancillary data
- Effects of cloud (masking/classification) and how to represent the effects to the user of data with adequate granularity but in an interpretable, ideally machine-readable manner

6.2 Pre-Flight Calibration of Satellite (and Airborne) Sensors

To ensure reliability of the measurement data from any sensor, it needs to be calibrated and characterised in a manner that accounts for the environment in which it is to be used. In this section, whilst we primarily discuss the needs for space, there are many similarities for other remote sensing platforms – particularly related to sensor size and effects of the operational environment. For space, pre-flight calibration in an operational environment means in a thermal vacuum chamber. Satellite sensors are calibrated pre-flight by organisations (industrial, academic and government research laboratories) that have suitable test facilities (clean rooms and thermal vacuum chambers). In some cases, instruments can be calibrated directly at NMIs; but, more commonly, NMIs provide reference standards for calibration in specialist facilities. In these cases, the NMI must be able to meet the highly demanding requirements for satellite calibration – often achieving state-of-the-art uncertainties under more challenging conditions (e.g. larger fields of view for radiometric sensors) and in cleanrooms. Providing NMI-quality calibration capability in a system that can be transported to the industrial calibrator's facility is particularly challenging, but very much in demand, too.

Although providing reference standards for calibration is a core traditional NMI function, the specialist nature of most of these transfer standards and the relatively small number of satellites/sensors requiring climate-quality uncertainties means that this only requires investment in capability by a small number of European NMIs. Therefore, the NMI community should seek to coordinate their efforts to ensure, where possible, that all satellite pre-flight calibration needs can be met, including the full spectral range for radiometric sensor (e.g. UV, visible (VIS), short-wave infrared (SWIR), thermal IR, microwave) and the calibration of components for active radar and LiDAR sensors.

For societal-benefit applications, such as climate, there is a recognition of the value of making pre-flight calibration information openly accessible to the scientists using the data. In our ECVs survey, some respondents complained about the lack of information about the calibration processes, writing for example:

> "Raw, uncorrected L1 data should be available to the user to ensure that potential errors in the correction methodology can be checked and does not need to be 'undone' if the user wants a different correction."

Satellite data taken now will have maximum benefit in future decades when long-term climate trends can be analysed. At that time, specific sensor expertise is likely to be no longer available, and therefore there is a strong need for the long-term preservation of not only data, but also of detailed documented information related to sensor characteristics which will be critical for its interpretation and trust. There is a role for metrology institutes to support the development of guidelines for this data and metadata preservation.

There are however, two barriers to providing detailed information about pre-flight calibrations that were identified in our satellite survey. First, the calibration methods are often commercially sensitive, with one respondent writing

> "The calibration is protected IP [Intellectual Property] and we will not disclose details of it."

Second, there was a concern that if the calibration methods were made public, the engineers and scientists involved in the calibration would be inundated with requests for more and more information and data that would be time consuming to resolve (and unpaid). Part of the problem is that historically the purpose of pre-flight calibration and characterisation was more to ensure that the sensor "met specification" rather than necessarily to determine its characteristics and uncertainty in-flight, given anticipated degradation. Calibration is also the last process in a long chain of processes from instrument design to launch and is often squeezed between a delayed manufacture and a fixed launch. There may be a role for metrology institutes in supporting the development of procedures for reporting calibration results in a way that respects the IP of the calibration institutes but provides information to the data users that is needed to define a full uncertainty analysis. More than ¾ of our respondents for the satellite survey said that NMIs had a role in supporting the development of rigorous uncertainty analyses for satellite sensor calibrations.

So far, most of the engagement of NMIs in satellite sensor calibration (pre- and post-launch) has been with passive radiometric sensors operating in the solar-reflective wavelengths (~300 nm – 2400 nm) and thermal IR (~3 µm to ~30 µm). In workshops such as the BIPM-WMO workshop of 2010 [87] and the Metrology for Climate Workshop [88], as well as in our survey, there is a common interest in having more metrological support for far IR, terahertz and microwave sensors – including active radar sensors as well as passive microwave. For example, in our survey there was a request for:

> "A spectrally-resolving, high-accuracy calibration source for the mid infrared,"

With the recent selection of several new satellite missions this need has further extended to the far IR (for FORUM) and includes characterisation of both detectors and the properties of materials. This demand for broad spectral range reinforces that of the BIPM-WMO workshop, which had as one of its key recommendations to:

> "Develop a consistent set of pre-launch measurements for microwave sounders for satellite agencies together with guidance to ensure SI traceability."

At the Metrology for Climate Workshop in 2015, there was a recognition that in this area:

> "There is some experience in the USA but Europe does not yet have consolidated expertise for MW [microwave] radiometry traceability."

There is still no microwave traceability in Europe [89]. The report of the 2015 workshop describes how the need is not just for the provision of suitable blackbody sources, but also for sub-system characterisation (antenna performance and detector linearity) and end-to-end characterisation (with a > 1 m diameter calibration target). The need for high-quality microwave calibrations is increasing in importance as the next generation of microwave sounders are looking more towards climate-focused observations for temperature and humidity trends, as opposed to the more immediate meteorological needs where short term consistency and resolution is perhaps a greater driver than rigorous SI-traceability.

6.3 In-Orbit Calibration and Traceability for Satellite Sensors

However well satellite sensors are calibrated pre-launch, the stress of launch and the harsh space environment degrades the instruments. Therefore, satellite sensors need to be calibrated or at least be monitored for changes in orbit. Such in-orbit calibration may involve onboard calibration reference standards that mimic those used in pre-flight calibration, although such onboard targets are also sensitive to degradation in orbit.

Thermal IR and microwave sensors are sensitive to their own operational temperature. Satellites with such sensors will typically use one or two onboard blackbodies as reference targets that are regularly (usually every few seconds) observed by the onboard sensors. As such sensors, and the satellites that carry them, get smaller, there is a need to design, develop, and characterise smaller, high-emissivity blackbody targets which places severe demands on knowledge of the properties of black coatings. As the demand to reduce uncertainty and evidence traceability become greater there are also concerns regarding the determination of temperature of such blackbodies and the reliability of the thermometer to indicate the emitting surface temperature, due in part to potential non-uniformities from the spacecraft and also change in thermal contact of the thermometer and degradation of its electronics. Thus, work is underway to integrate transition cells (metal or eutectic freezing/melting points) into the blackbody as SI traceable fixed-point references.

In the larger, public-space-agency operated satellites, VIS and SWIR sensors typically use white panel diffusers to reflect sunlight into the input optics as a source of nominally calculable radiance. Some satellites have carried two such diffusers so that one can be used regularly and the other occasionally to monitor degradation of the operational diffuser in orbit. Such comparisons show that these diffusers can age considerably in space, particularly in the shortest wavelengths, making it difficult to have confidence in the performance of the system that in principle is seeking to provide the calibration reference in space, or at least at the uncertainty levels required for climate [90], [91]. One of our survey respondents wrote:

> "In some cases, there is insufficient reference data to truly capture the time-dependent calibration coefficients, particularly in the UV. Also, some calibration coefficients drift over the course of an orbit, every orbit. It is hard to parameterise this drift over an orbit if you have only the Earth radiance as input, because of the changing scene."

Very few sensors have flown sources to characterise their spectral response functions in space, and the in-orbit degradation of spectral response function for radiometric sensors, particularly those constructed from physical spectral filters as opposed to spectrometers, is not well understood. One response from the survey stated a need to:

> "Separate the time-dependent bias correction values of the spectral signal and the time-dependent correction of the instrument spectral response function."

One of the ways that in-orbit degradation can be evaluated and corrected or validated, is through the observation of natural and artificial Earth reference targets. For active radar sensors, point-source emitters and corner-cube reflectors are used, along with artificial reflectance targets that test instrument parameters. For passive sensors in the solar-reflective spectral region, a broad range of natural and artificial sites are used for both geometric and radiometric cal/val. These include the FRM sites where ground measurements can be compared to the satellite observations using field instruments and methods that have some assessment of uncertainty, for example the CEOS RadCalNet sites [92]. However, even if this starts to help in addressing challenges with pre-flight and on-board calibration, as a survey respondent explained:

> "One "barrier" which is currently being worked on: progress is being made towards the inclusion of vicarious calibration sources (RadCalNet) in the calibration workflow (previously: on-board cal. with life-limited items)".

Achievable uncertainties are limited to a few percent at best, and not adequate for the most demanding requirements of climate. Many deserts and targets like the Moon have good stability but limited knowledge of absolute radiometric uncertainty. In some cases, particularly for thermal IR sensors and for visible sensors operating over the oceans, the signal measured comes in almost equal proportions from the ground and from the atmosphere. In these cases, the ground observations cannot be directly compared to the satellite sensor data, but only to the product derived from the satellite sensor data, which has been processed through an algorithm that corrects for the atmospheric scattering and emission (including a radiative transfer code). Similarly, for vegetated scenes, there are significant issues of scaling between the observation made by the ground-based or aerial instrument and that seen by the satellite. With active sensors too, the only option is to compare the derived product within situ measurements.

For traditional applications of satellite data, cal/val processes have been developed and refined by the communities over several decades. The enhanced uncertainty requirements required for climate studies mean that these methods are being rethought at present, especially in terms of developing the FRM networks. As one of our survey respondents wrote:

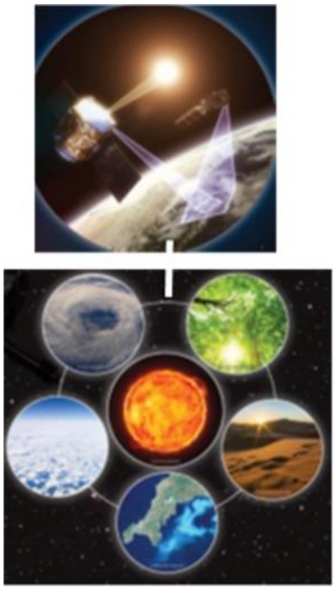
> “Level 1 data need to be improved to be within < 1 % absolutely.”

A shift from the traditional 3 % - 5 % uncertainty level to sub 1 % requires more metrological analysis. There are roles for metrology in providing SI-traceability to these ground observations, in developing the data science methods to combine data from multiple comparisons to obtain additional parameters (e.g. to see spectral response function changes from comparisons over different surfaces), to develop methods for scaling data, and to support the formalisation of such comparison processes.

The final method for post-launch cal/val is comparison with other satellites. Such comparisons may take place using “matchups” (when the satellites see approximately the same scene at approximately the same time), or using “pseudo-invariant” sites, where the site is relatively stable between different overpasses, or through using instrumented ground sites where differences between the two overpasses are measured and can be corrected for. Usually, because of orbital dynamics, near-simultaneous matchups occur only in a very small number of locations, usually around the Poles, although TRUTHS (see case study) will intentionally fly on a different orbit with more frequent matchups. For active sensors where it is essential to have long-term consistency and bias removal to detect small trends, there is often an intentional “tandem” period at the start of a mission where a satellite flies in close proximity to another sensor for a period of several months so that they can be cross-compared across the whole globe, since reliance on SI-traceability of the sensor itself is currently inadequate. Most recently, Sentinel 3B flew just 30 seconds ahead of Sentinel 3A for three months shortly after launch, before being drifted into its final orbit. As the two Sentinel 3 satellites (with identical instruments) carry not only active sensors, but also both visible and thermal IR passive sensors, this also gave an opportunity to do a tandem comparison between passive sensors, illustrating, in the absence of robust SI-traceability in flight, the benefits and arguably necessity of such an approach. Metrology institutes have engaged with the science teams performing such inter-sensor comparisons, but there is still a lot of work to be done to bring a full metrological rigour to such comparisons, and to provide tools to handle uncertainties, error covariance and the combination of many comparisons across a wide range of sites. These comparisons are particularly challenging data science problems (typically with 100s of millions of matchups and error correlations to consider).

Of course, the best comparisons would occur if SI-traceability were brought to orbit. There are a few “SI-calibration satellites” in development, including CLARREO by NASA and CRAB by the Chinese Space Agency. The TRUTHS mission, conceived and led by NPL, funded predominantly by the UKSA, is now under development by ESA. This, and the sister missions, in effect seek to establish a very high accuracy reference in space that is SI-traceable, delivering not only its own data but also serving as an in-flight calibration reference to transfer its accuracy to other satellite sensors, in effect taking the ‘NMI into orbit’.

Case Study: TRUTHS



The need for SI-traceable uncertainty for space-based climate observations is now well-established [78], [79], [81]–[96], as is the inability of the existing space observing system to achieve the uncertainties required for detection of decadal scale climate trends. At present, it is not possible to provide robust evidence of SI-traceability on-board any spacecraft. The TRUTHS mission was conceived at an NMI (NPL) nearly 20 years ago, with the aim of establishing a satellite to make full SI-traceable measurements in space of the Earth and Sun at an uncertainty level commensurate with the needs of climate.

The mission is now in development at ESA and its disruptive innovation is to mimic the terrestrial traceability chain of an optical imaging sensor including the flight of the primary standard – a cryogenic radiometer – in space. The mission concept is described in more detail in [97] and not only measures the climate state of the planet but also upgrades the performance of other satellites through reference calibration using simultaneous overpasses (its orbit is established to provide more frequent matchups at every latitude), as well as through its measurement of pseudo-invariant and instrumented calibration sites.

The acceptance of a “metrology mission” demonstrates the increasing recognition of the importance of metrology to EO and climate, and shows that meeting the stability and consistency goals needed to identify a climate record benefits from ambitious, innovative solutions, alongside a strong partnership between experts from the whole community – academia, industrial engineering and NMIs.

6.4 Ground-Based and Airborne Remote Measurements and Networks

Complementary to satellite measurements are measurements performed in ground-based networks like AERONET, BSRN, the NDMC, or by airborne sensors operated from research aircrafts/balloons, etc. Even though such networks do not get the full global coverage achieved with satellite observations, worldwide-distributed networks provide a pointwise global view with local measurements of possibly very high quality, and a good height resolution in the atmosphere. Similar to the FRM concept, discussed earlier in Section 6.3, the combination of satellite and ground-based measurement networks or airborne measurements enables mutual redundancy and consistency.

The advantage of the ground-based sensors is that they are often easily accessible and could, in principle, be recalibrated on a regular basis. One difficulty is operating tens or even hundreds of sensors in a compatible manner and ensuring traceability to a common reference. This implies different concepts to realise the traceability of those systems, provide recalibration on a regular basis, and common methods to assess the uncertainties of these systems. These requirements are underpinned by the writing of a BSRN member in the survey:

> “As member of BSRN, I should aim to reach BSRN-defined accuracy targets for the radiation data I generate. Demonstrating such targets are fulfilled requires computation of **operational** uncertainty for all produced ECVs. This requires an agreed-on determination of the uncertainty following guidelines of GUM. Systematic research establishing such uncertainty for all radiation ECV is necessary.”

Similar concepts to those already successfully applied in the past within the MetEOC projects would be required to establish traceability for a few stations within the NDMC network, consisting of a combination of travelling reference sources and travelling reference detectors for redundancy. Sets of these in combination with intercomparisons are a way to establish traceability for complete networks. For the BSRN and its reference sites, groups of instruments have been established to provide a stable reference, with the goal of SI traceability but the requirement for stability. Efforts to reduce the dependency on these artefact-based scales to the robustness of SI is a focus of the community, and one highlighted by the WMO in signing the MOU with BIPM. One such group, the WISG, not only requires this SI traceability but also a reduction of the measurement (standard) uncertainty from 10 W m^{-2} to 2 W m^{-2}. Efforts to achieve this are happening through projects like MetEOC which is developing a combination of a dedicated reference blackbody and radiation thermometer.

New types of sensors are often developed for ground-based or airborne measurements first to allow for continuous improvement and fine tuning in order to achieve the maximum performance, before being incorporated into satellite experiments. An example for an airborne sensor is the GLORIA instrument on the German research aircraft, HALO. GLORIA consists of a hyperspectral camera operating in the mid infrared (MIR) and in combination with specific flight patterns even allows the detailed tomographic investigation of three-dimensional objects like thunderstorms. Here a dedicated traceability concept was developed within MetEOC and successfully applied. Examples for other new developments are sensitive MIR and far infrared (FIR) radiometers requiring tuneable MIR sources for characterisation:

> “I am working towards new sensors in infrared limb sounding. One of the main problems is the interaction of different error sources. A good radiance calibration and provision of high-quality calibration sources (blackbody, pre-flight and in-flight) is a huge asset, which confines one of the major error sources. If in future new filter radiometers for the mid IR (<15 µm) should play a larger role, an accurate radiometric spectrally resolved calibration source would be needed for end-to-end calibrations of instruments”,

or rugged spatial heterodyne sensors requiring reference sources with well-defined wavefront curvature:

> “We work on a particular spectrometer technology called 'Spatial Heterodyne Interferometer' suited for nano- and micro satellites. The metrology to characterize such instruments is not established at all. Such instruments' data is not used for science applications so far, to our knowledge. But it is a promising technology.”

These examples illustrate the need for continuous exchange with the community and support with the development of appropriate reference sources and concepts.

6.5 Retrieval Models

There are many situations where the direct measurand observed by a sensor needs some form of model or algorithm to enable the bio-geophysical parameter of interest to be extracted or retrieved. Whilst it is not necessary for the metrology community to be expert in developing the theory underpinning the transformations, metrologists are looked to for guidance on propagating uncertainties through such processes and on how to report resulting uncertainties. In some cases, the non-linear nature and need for data analytical tools and methods such as neural networks can add a further challenge, as can the volume of the data. This challenge of course pervades across all the themes in this EMN.

In satellite remote sensing, when the observed top-of-atmosphere (TOA) radiances are known instead of the full atmospheric and surface state, inverse radiative transfer modelling can be applied to retrieve the best fitting ECV values. Optimal Estimation (OE) techniques are often used to do this kind of retrieval. The same considerations that apply for surface reflectance retrieval also apply to the retrieval of ECVs. This suggests [80] that the metrology community can help in understanding the sources of errors in RTMs used by many ECV products in the retrieval process, thus supporting data providers and the research community in developing good practice guidance. There are often several retrieval algorithms used by different groups to derive climate data even when the input data is the same. In order to develop an optimum algorithm which provides the best estimate, more cross comparisons are needed to ensure that any given retrieval is as good as it can be, for example, undertaking round-robin exercises to understand the relative performance and strengths of different methods. The metrology community can support this process by helping to organise and report results of such comparisons and validating the algorithm's performances in order to establish its degree of consistency with reality.

For atmospheric composition measurements from satellites, there is also a metrological need associated with the retrieval of atmospheric parameters due to the degenerate nature of these retrievals. Many atmospheric retrievals deal with ill-posed problems, meaning at least some part of the retrieval has too many free parameters to be constrained with the available information. Therefore, either regularisation is applied [98] or priors are constructed and imposed [99]. However, the uncertainties associated with regularised estimators are very difficult to constrain. In addition, there is only limited knowledge on what priors are appropriate for the various parameters. However, here it does need to be noted that the strongest meteorological need is for the most degenerate parameters (e.g. vertical profile of CH_4), yet there are also other parameters that are better constrained (e.g. total column CH_4). Regardless, due to the large number of different parameters going into these retrievals, a systematic study of their uncertainty contributions to the various measurands would be beneficial to the field. The EU H2020 project GAIA-CLIM [50] highlighted many of the challenges facing atmospheric composition retrieval.

For land and ocean applications, instead of concern as to how atmospheric composition influences the atmospheric radiation budget and drives climate change, we are instead concerned with how it impacts the transmission of the desired ground signal to the sensor. Molecules and aerosols in the atmosphere absorb and scatter photons, thereby affecting the light path and the resulting radiation field arriving at the satellite. Atmospheric RTMs are used to enable atmospheric correction to be applied to the satellite data of interest for a given set of predetermined atmospheric parameters (such as the distribution of the various molecules and aerosols). This allows us to compare TOA and surface observations, which makes RTM a key step for the in-orbit calibration and validation of remote-sensing radiances observed by satellites.

However, there is of course an uncertainty associated with this atmospheric correction. The state of the atmosphere is never known perfectly, and the uncertainties on each of the input atmospheric parameters will propagate through to the measurand. In addition, there are model uncertainties to account for errors in the assumptions made within the RTM and numerical errors. Govaerts [100] found differences up to 3 % between various RTMs using the same set of atmospheric parameters. The validation of future in-orbit climate missions will require better accuracy than this 3 %. Improving the accuracy and traceability of RTMs is thus an important goal. The Eradiate RTM [101] is currently being developed in an effort towards achieving these goals and as an open-source European RTM that is strongly supported by ESA and the European Commission. By combining Eradiate with metrological satellite missions (e.g. TRUTHS or CLARREO), it will be possible to bring SI-traceability to RTM, thereby closing the traceability-loop between the TOA radiances, surface reflectance, and the radiative transfer calculations in between. Although there is an urgent priority for the solar reflective domain, where the effects are most complex, similar activities will be needed to extend these improvements into the thermal IR spectral region also.

In addition to the functionality and physical description of the RTM code, the parameterisation of the atmospheric composition and associated scattering/absorption cross-sections of the molecules and particulates also need to be as correct as possible. Thus, the development of robust spectroscopic libraries to enable representative parameterisation of the models is also a need, although one that is addressed within the atmospheric theme of this EMN, Section 3.5.

RTMs are not only important for the calibration and validation of Level 1 data, but are also a tool used in various EO fields, as they can be used for the propagation of optical photons through environments other than a cloudless atmosphere. For example, photons can propagate through thin clouds or surface snow, ice or water; or the interactions between light and a forest canopy can be modelled. These various environments each have different requirements and different RTM are thus used. Typically, either 1D or 3D simulations are used, using either the discrete ordinate method or ray-tracing Monte Carlo techniques. In general, there is a wide variety of RTM, each with their own set of uncertainties. Here we are talking about “forward modelling”, meaning the input parameters (e.g. the state of the atmosphere) are known, and used to calculate how light propagates through this environment.

If instead the final radiation field is known (through observing it with a well-calibrated satellite), RTM can be used to get constraints on one or more of the input parameters. In this case, many forward model RTMs are run with various values for the unknown input parameters, and the best values can then be determined by comparing the resulting radiances from each forward model to the observed satellite radiances. This kind of `inverse modelling’ is used for the retrieval of atmospheric properties and surface reflectance.

Surface reflectance modelling usually requires the simultaneous fitting of many free parameters. The CEOS Atmospheric Correction Intercomparison Exercise (ACIX) [102] compared the reference data set with some standard surface reflectance products, providing information on product performance. Due to the large number of free parameters, surface reflectance modelling requires computationally expensive techniques, which complicate the uncertainty propagation to surface reflectance, as robust techniques like traditional Monte Carlo become unfeasible. One of our survey respondents wrote:

> “Having the absolute and relative radiometric calibration uncertainty at TOA [top-of-atmosphere] Level (L1B) is nice (and required for BOA [bottom-of-atmosphere]), but for the "end user", the uncertainty in BOA reflectance (per observation or per pixel) is the critical part still missing.”

Optimal Estimation (OE) techniques that are regularly used for surface reflectance modelling do allow for the accounting and propagation of uncertainties [103]. The main difficulty consists of determining appropriate surface and atmospheric priors, as well as covariance matrices for both the measurement uncertainty and model unknowns. OE uses Jacobian matrices that result in locally-linear approximations of the probability density [104]. The results and associated uncertainties are thus only valid for a very limited range of surface and atmospheric states. Markov Chain Monte Carlo (MCMC) methods allow the sampling of more complex probability density functions, which include bias, local minima, and correlations in state estimates [105]. These methods are valid for a wider range of states, but there is still a need to test and validate the results, especially in challenging atmospheres with high water vapour and aerosol loadings [106]. The non-linearity and inter-dependence of some of these retrieval algorithms means that researchers are looking to neural networks and machine learning techniques to provide optimal solutions, but at present uncertainty propagation through these methods remains a challenge.

6.6 Summary of needs for Remote Sensing from metrology

Metrology Challenges for Remote Sensing
Pre-flight calibration standards and methods to enable SI-traceable uncertainties that commensurate with the needs of climate, available for cost/time efficient calibrations under operational conditions at industry/academic locations. These should cover the needs of all sensor domains e.g. passive optical through to active microwave and address all necessary parameters and observational platforms e.g. space through to surface networks.
On-board calibration standards and methods to enable SI-traceable uncertainties that commensurate with the needs of climate to be achieved. This requires NMIs to support the transition of terrestrial techniques to orbit, assess degradation and uncertainty estimates for their use for all technology domains.
Establishing FRM quality test-sites/measurements for post-launch Cal/Val this includes enabling SI-traceability in the field, uncertainty evaluation of in situ and its propagation/representativeness to the satellite all domains.
Development of a metrologically-based QA framework and associated methods/tools to facilitate evaluation and consistent reporting of end-to-end uncertainty of Level 1 and consequential higher-level data products. Including training support on uncertainty evaluation.
Metrological assessment of uncertainty of models & algorithms particularly those required to transform top-of-atmosphere measurands to bottom-of-atmosphere parameters, including support to developers, guidance on assessment, such as the challenges of ‘machine learning’ methods. Means to establish uncertainty characterisation and representation for ‘classification’ systems e.g. land cover type, cloud masks etc. is also needed.
Comparisons and guidance for organisation of comparisons of community ‘measuring systems’. Comparison of satellite-to-satellite and satellite-to-ground systems following metrological best practice to establish ‘degrees of equivalence’ to enable international interoperability and harmonisation for long time-base FDRs.
Methods to establish metrologically robust FDRs and CDRs. How to combine both similar (e.g. sat series) and differing (different sensor designs) sources of data to create long-time

base (multi-decadal) series with associated uncertainties. Facilitating interoperability (bias removal).

7 GENERAL CROSS-CUTTING METROLOGICAL SUPPORT FOR CLIMATE AND OCEAN OBSERVATION

In previous sections of this report, we have considered the needs of specific communities. Here, we consider overarching trends and needs that are relevant to multiple applications as those relating to cross-cutting themes.

7.1 A Metrological Approach to Modelling and Reanalysis

In climate and ocean observation, modelling is crucial. Models are used to interpret observations, compare different types of observation, transform measurands to bio-geo-physical parameters, bring observations into reanalyses and test climate prediction models.

Models are used within observations themselves. For example, a satellite observation of sea ice volume will involve an altimeter signal, which is corrected according to how much the atmosphere has slowed down the radar pulse (involving models of atmospheric conditions), along with an estimate of what proportion of the iceberg is under the water (based on models of ocean and ice density, temperatures, and estimates of snowfall on top of the ice). For in situ observations the measurement may be more direct and specific, but the interpretation and scaling to a gridded product, more representative of the real environment, involves modelling. Modelling is also essential to compare and combine observations. For example, when the ECV sea surface temperature is observed in situ by an ocean buoy, it might be observed by a thermometer about 20 cm below the surface of the water. Meanwhile, a satellite observes a thermal IR signal that comes both from the top few microns of the water and from the atmosphere. In order to remove the atmospheric component from the observation, an atmospheric radiative transfer model is used. In order to compare the satellite and in situ data, a model for temperature variability in the top few centimetres of the water is needed (and this, in turn, depends on wind speed, solar irradiance and ocean salinity).

Climate scientists, meteorologists and oceanographers use data from observations and their uncertainties in modelling processes. Climate models themselves do not use observational data directly in their prediction models, but such models are run as a “hindcast” (starting in the past) so that the output of the model can be compared with observational data. Climate scientists rely on such comparisons to understand properties of their models. Observational data is, however, used directly in reanalyses (historical climate records) and for numerical weather predictions (near real time predictions), though the process of “data assimilation”. Data assimilation uses the difference between models and measurements to inform model parameters. To provide the right weighting to such differences, and to understand potential biases, it is necessary for the observations to be provided with robust uncertainty analyses, as well as information about error covariance structures. The observational uncertainties include both measurement uncertainty and uncertainties associated with the representativeness of the measurement to the model cell and the uncertainty associated with any radiative transfer processes. Data assimilation techniques are generally set up to use an error covariance matrix for the observations, which are required for correct parameter determination, but observations are rarely provided with robust covariance information.; As a result, it is inferred from diagnostics from the data assimilation system. With neither error covariances, nor observational uncertainties, fully understood, data assimilation for reanalysis does not generally rely on the uncertainty assessments of the observations, aside from for a small number of “anchor observations” (currently entirely in situ data sets). This is because these are not considered sufficiently robust. The assimilation process adjusts both data and model to produce a fit, including “bias correcting” measured data by more than the declared uncertainties, with all the risks that can entail, such as missing real phenomena that are unexpected or very local.

Reanalyses are models of the historical global climate (atmospheric, ocean and land parameters) that are determined by assimilating all available observational data into a global climate NWP model which has a fixed configuration. For example, the ERA5 reanalysis, developed by the ECMWF assimilated over 87 billion observations to provide hourly global data on a 31 km grid from 1950 to within 3 months of real time (updated continuously). The outputs of such reanalyses are considered observational datasets and are available on the Copernicus website. Furthermore, they are used by observation networks and satellite data processors as inputs to corrections applied to measurement data.

At the FIDUCEO project workshop in 2019, William (Bill) Bell of ECMWF explained how the reanalysis team needs more robust uncertainty information, and particularly information on systematic effects and error correlation structures, in observational data.

Data reanalysis models do not generally provide usable uncertainties. However, ERA5 is the first reanalysis that provides an ensemble of values to users, so that users can run multiple examples through their own processing chains to understand the effects of possible model uncertainties. The ensemble spread is based almost exclusively on the uncertainty associated with random effects in the observations and some known uncertainties in model parameters. Bell noted that work was being undertaken to ensure more robust uncertainty information was available in future reanalyses and specifically requested the involvement of the metrology community. He said:

> “We are currently looking at sea captain records from the 1850s and we wish we had a time machine to go back and ask them to record more details about how they did their measurements. What will scientists of 2100 wish we had done? – Provide more detailed and more robust uncertainty information.”

Merchant *et al* [107] also discuss the importance of rigorous uncertainty information in reanalyses. In our survey, one response described the value of metrological research in reanalysis products:

> “Yes, metrological research would improve the error characteristics of the mean state of ECVs from reanalysis products.”

Finally, machine-learning techniques, including neural networks, are increasingly being used in the generation of ECV products, particularly for more complex variables that rely on proxy measures or to associate satellite observations with an in situ metric. Neural networks are also used in applications of ECVs beyond direct climate modelling, and particularly in “climate services”, for example to establish risk or to quantify “embedded carbon”.

Metrologists, and particularly data scientists working in metrology institutes, can support the analysis of uncertainty through modelling and the interplay between measurements, observations and models. Metrological research is needed to understand how model uncertainty is evaluated and validated, to develop standardised ways of reporting results of model performance evaluation, and to provide methods for uncertainty propagation through non-linear models. Furthermore, error covariance information is very important, but with huge data sets (87 billion observations) new methods are needed to process such information in a computationally affordable manner, and methods such as machine learning will themselves need to be robustly evaluated to assess the uncertainties they might introduce into the analysis through their use.

7.2 Low-Cost Sensors

Low-cost sensors are here defined as those which are, at most, one tenth of the cost of the reference sensors typically used by a community for environmental observations. They are usually smaller in size, mass and power consumption and are increasingly being developed as miniaturised components become available. Here we include both low-cost sensors developed for meteorological or air quality purposes, and commercial satellites, which are very much cheaper and much smaller than the satellites launched by national and international space agencies.

Low-cost sensors have several advantages over more accurate, but bulkier and more expensive, “reference instruments”. Most significantly, there will be many more of them. This means that they can provide local information, filling in detailed spatial information between the reference observations. Air quality and meteorological sensors can be used to make localised information within cities where city structures strongly influence local conditions, methane sensors can be attached to cows to improve understanding of agricultural GHG emissions, thermometers attached to surfboards to obtain a measure of sea surface temperature in the difficult-to-observe coastal waters [108], and constellations of commercial satellites can provide very high spatial resolution data with hourly temporal coverage [109]. Low-cost sensors can also provide data access and societal engagement from a wider community through “citizen science” and enthusiastic amateur observations (e.g. hobbyists having weather stations in their own gardens), through personal exposure observations (e.g. of air quality) or bespoke instruments providing valuable local information for e.g. for a farmer, small scale fishing enterprise or commercial enterprise. Data collected for such immediate and local applications can also have longer term value for climate science and oceanography, providing its uncertainty (sensor and usage) can be appropriately evaluated. Similarly, commercial satellite data, valuable in the short term for commercial and social applications (disaster monitoring, fishing, agriculture, city planning) may also have a longer-term value again providing the data quality can be quantified.

In 2018, the WMO published a review on the performance and applications of low-cost atmospheric composition sensors [110]. According to this review, the currently available technologies (sensors) are susceptible to variable and unpredictable behaviour. These sensors typically suffer from poor stability/reproducibility, low accuracy and exhibit high cross-sensitivities to compounds other than the target species. Furthermore, a reliable metrological validation and a robust characterisation of the measurement uncertainties are usually missing. However, the review concludes by describing three (increasingly valuable) approaches to characterising these sensors so that they can be more reliable and their data usable:

1. comparison of their data with nearby data and/or models,
2. colocation of low-cost sensors with higher quality instruments for direct comparison data, and
3. rigorous laboratory calibration of the low-cost sensors at regular intervals using a quality-assured calibration process.

Metrology could provide support for improving the quality of all three levels of characterisation and to help users determine the fitness for purpose of their calibration/comparison scheme, although the adoption of the steps needed for the increased rigour may increase the cost to a level where the benefit is reduced too much and so strategies to keep costs low are needed.

At the stakeholder needs workshop held by this EMN in February 2020, Bertrand Calpini of the WMO emphasised the growing importance of low-cost sensors and crowd-sourced meteorological data. He would like further collaboration with metrology in developing guidance on how such sensors could be calibrated and their data quality controlled. He also emphasised the importance of judging a sensor according to its “fitness-for-purpose” for a specific application.

This recognition of “fitness-for-purpose” is also important for commercial satellite sensor data. In satellite EO there is currently an effort to define “analysis ready data” as a concept that allows users to access data more easily and to assess its fitness for their purpose in a user-friendly manner. Such discussions benefit from the contribution of metrologists in defining an appropriate QA framework.

7.3 Historical data and Data Rescue

Much of the focus of this report has been in establishing improved SI-traceability and robust metrological analysis of current observations. An equally important aspect is to collate, quality assure and provide recalibration and uncertainty analysis on historical data, particularly data that has been originally collected for other purposes. For example, ships’ water temperature logs can provide information on sea surface temperature into the 19th century, early meteorological data is also extremely valuable and early satellite data from the 1980s (and even 1970s in some cases) can provide global coverage over much longer timescales.

The WMO Commission for Climatology (WMO-CCI) is coordinating the reuse and preservation of historical in-situ data through a specialist Data Rescue Expert Team. Their Guidelines on Best Practice in Data Rescue [111] focus on the copying of data records and QA to remove miscopies and outliers. Elsewhere, researchers are analysing how to interpret historical data. For example, in [112], researchers evaluated the warm bias on European meteorological data collected before screens were introduced. Metrologists and data scientists working in metrology institutes have been involved in such projects, and there is considerable value in further engagement.

Early satellite data provides another extremely valuable data source due to the global coverage of low-Earth orbit satellites and time series over wide areas for geostationary satellites. However, early satellites were calibrated to show that they met their specifications for short term applications, and metrologically-rigorous information about pre-flight calibrations are not available. Furthermore, all satellites change in orbit and those processes were not always well monitored in early satellites. In many cases there is limited information available because of limited downlink capabilities and data is archived on tape stores and therefore not readily accessible. Many of the world’s space agencies are running “long term data preservation” programmes to collate and make accessible such historic data, as well as to provide FCDRs that archive the necessary metadata and auxiliary information of current sensors for future users. As part of such programmes, data are often reprocessed in new “collections”. This reprocessing is an opportunity to improve the metrological rigour of data sets obtained from sensors. The FIDUCEO project developed a methodology [75] for performing and documenting a metrologically-rigorous uncertainty analysis for historical sensors as part of reprocessing raw (Level 0) data to an FCDR. Such methods were also adapted by the GAIA-CLIM project for in situ data [50]. FIDUCEO also used “match ups” (overlaps between satellites) to recalibrate the historical satellites with respect to modern satellites. The GCOS implementation plan [113] states:

> “The priority with regard to the management of early satellite data is to ensure long-term preservation of the raw data and Level 1 data for input to FCDR production. Progress towards preservation of historical satellite data has been made both for geostationary and polar-orbiting meteorological satellites, but the associated critical metadata are more difficult to preserve.“

The potential value to extend backwards in time the start of a measurement record is significant considering in general it takes many decades for any clear signal of climate change to be of sufficient size to be detectable from a background of natural variability compounded by uncertainties and drifts in measurement instrumentation and the need to link datasets robustly in a harmonised manner. However, whilst the benefits of long-term records, including data from early sensors are clear, they are only valuable if an appropriate level of trust can be placed on the value of any historical data set and that it can be reliably linked with that of current and future sensors and have associated with it a reliable estimate of uncertainty.

In many cases, particularly where satellites were part of operational observation programmes typical of meteorology, efforts were often made to overlap observations for a period of time with a view to normalise the data outputs i.e. removal of bias. Whilst this process was beneficial, it has some limitations in terms of climate:

- Overlap time may coincide with an abnormal climate event e.g. El-Niño or volcanic eruption
- Subtle differences in sensor characteristics, e.g. spectral shape of bandpass filters, can introduce anomalies dependent on scenes used for harmonisation
- Observation of common scenes may not be simultaneous or identical in nature
    - Common scenes may not have adequate temporal stability and may not be well-defined for some types of sensor
- ‘Reference sensor’ may have performance drifts between time it was calibrated (harmonised with previous sensor) and its use with follow-on sensor.

A detailed ‘forensic’ approach following the principles established in FIDUCEO needs to be pursued for the full range of sensor types. Such analysis is challenging and requires not only a detailed understanding of the sensor technologies used today and previously, but also an understanding of the principal sources of uncertainty, interdependencies and likely performance decades previously. Metrologists have a role to play, not least in helping sensor experts in their analysis and sometimes with knowledge gained from terrestrial applications in other fields, but also in establishing common frameworks to allow reporting in a similar manner, facilitating interpretation by current and future scientists.

7.4 Paleoclimatology

Paleoclimatology is the study of climates before measurements were taken. Paleoclimatology uses proxies to evaluate earlier states of the climate using rocks, ice sheets, tree rings, corals, shells, microfossils and similar techniques, and can be used to estimate climatic conditions from a few hundreds of years until a few hundreds of millions of years before present. Such analyses can provide valuable information on the natural variability and evolution of the Earth’s climate system. They are particularly interesting for understanding the key triggers for abrupt, irreversible climate changes (tipping points).

Paleoclimatology is based on measurements of samples collected on field campaigns. Models are used to connect the proxy observations to the climate variables of interest. Many of the challenges in paleoclimatology are similar to challenges in modern climate science – understanding uncertainties and error correlation structures for the measured parameters, and understanding limits to the models. A paleoclimatologist we spoke to in preparing this report told us that uncertainty analysis is particularly unreliable in this field, which does not have the same formal structures and methodologies as instrumental measurements. The unreliable uncertainty estimates make it difficult for models to be constrained. She said:

> "Expeditions to collect ice cores from remote locations are expensive and yet the data they provide are often unreliable. Improved metadata, more consistent sampling methods and more rigorous uncertainty analysis (including error correlation information) would allow that data to be used more reliably in paleoclimatological models."

7.5 Uncertainty Analysis

The most common topic arising in our survey, as well as in community strategies and workshops, was that robust uncertainty analyses are needed for climate and ocean observations and that the metrology community can support the development of uncertainty budgets. For example:

> "Metrology institutes could help scientific communities to understand the breadth of error sources in their data and how these can be quantified in uncertainty estimates."

It is important that such uncertainty analysis is not just limited to the instrument or its calibration, but to the use of that instrument under field conditions:

> "As member of BSRN, I should aim to reach BSRN-defined accuracy targets for the radiation date I generate. Demonstrating such targets are fulfilled requires computation of **operational** uncertainty for all produced ECVs. This requires an agreed-on determination of the uncertainty following guidelines of GUM. Systematic research establishing such uncertainty for all radiation ECV is necessary."

As well as providing uncertainties under operational conditions, it is also necessary to understand the error correlation structures between measured values obtained at different times, or obtained from different instruments in a network, or, for example, in different spectral bands for a radiometric quantity:

> "Metrologists should help develop a] better treatment of distinction between systematic and random uncertainty components."

This is something that many experts in the community find difficult and perhaps do not have time to do (uncertainty analysis is, perhaps surprisingly, not always considered a valuable use of a researcher's time):

> "Yes. [Metrologists can help with the] analysis of error correlation structures, which is slow, difficult work that doesn't usually result in publications and so can't be done by postdocs without some amount of self-sacrifice."

Support with uncertainty analysis is needed for both satellite and non-satellite measurement techniques, for instruments that make reasonably direct measurements of an ECV quantity (e.g. near surface air temperature), to observations that are processed through extremely complex processing chains, including models and even neural networks (e.g. leaf area index).

> "A grounded framework for the propagation of uncertainties in complex algorithms is still missing in the field."

Such support is requested for physical and chemical quantities, and an interest in whether and how biological ECVs and EOVs can also have quantified uncertainty statements.

Uncertainty support that is requested falls into three main categories:

- Development of general frameworks / approaches to perform uncertainty assessment for a community (set of ECVs);
- Direct support from metrology organisations to help with a specific uncertainty budget for a particular measurement or ECV process;
- Development of tools to simplify uncertainty analyses and its reporting.

The C3S “Lot 3.11” invitation to tender included a requirement to develop a:

> “Software tool for evaluating uncertainties in near surface temperature measurements, based on metadata and completeness level.”

7.6 Vocabulary and Ontology Definitions

Alongside requests for support with uncertainty analysis, the second most common request in our survey and workshops was for support with the standardisation of a metrological vocabulary for observations. In many cases, papers and reports by the different communities involved in climate and ocean observations, confuse the terms “error” and “uncertainty” and are inconsistent in the use of terms such as “bias”, “noise”, “random and systematic” etc. Furthermore, some communities are driving for the definition of conceptual terms such as “interoperability”, “analysis ready data” and “harmonisation”, with the expectation that such definitions would define a process and standardisation, as well as a word.

In [80], the first recommendation was for a standardised metrological vocabulary:

> “The metrology community should provide a standardised ECV vocabulary consistent with those being defined through international coordination bodies for satellite and in situ observations (i.e. CEOS, GEO, ICOS, IPCC), in order to avoid the misuse of some words such as ‘errors’ and ‘uncertainty’. Standardised definitions of all terms related to measurement can be found in the International Vocabulary of Metrology [VIM]. The vocabulary should include examples as well.”

Similarly, the WMO-GAW implementation plan [12] states

> “A-QA-5. Adopt and use internationally accepted methods and vocabulary to quantify the uncertainty in measurements. To promote the use of common terminology, a web-based glossary was developed.”

Increasingly, there is a move towards increased machine-readability of climate data sets, with defined metadata and ontologies (hierarchical computer-readable vocabularies). For example, the NetCDF standard [114] is widely used for the storage of large arrays of observational data (particularly, but not exclusively, from satellites). NetCDF is supported by Climate and Forecast (CF)-Convention standard names [115]. These standard name lists have multiple names for any particular measured quantity, but there are no defined terms for uncertainty in most ontologies. Where such terms exist, they are limited and do not have sufficient flexibility to break down uncertainties, e.g. to distinguish standard or expanded uncertainty, or to describe uncertainty associated with common or independent effects.

There is a need for metrology, and particularly data scientists and curators working within NMIs, to improve the ontology standardisation to include sufficient information about uncertainties to store complex error covariance structures in time series and spatial data.

7.7 Training in Metrological Techniques

Another common response to our survey and request in workshops is for training in metrological techniques. In our survey one respondent simply responded:

> “We need more training.”

Often, the request for training was around uncertainty evaluation. The WMO-CIMO Expert Team on Metrology have as their first item on their work plan to:

> “Develop training modules on uncertainty calculations”.

While generic introductory training was valuable:

> “Please keep up the nice availability of "entry-level" reference documents (guides / tutorials related to the GUM),”

Most communities want training that is tailored to their specific applications. ESA and CEOS have asked for further development of the existing NPL training courses on uncertainty analysis for EO so that they cover a wider range of satellite types (active as well as passive sensors) and are extended to cover the processing of ECV quantities from satellite data through complex, multi-step processing chains.

There were also requests for practical training in measurement best practice and instrument calibration. WMO-CIMO has requested metrological training for extreme environments and key climate regions, as well as training on the principles behind measurement techniques (including quantity definitions). Some communities requested training in establishing ISO 17025 standardisation for calibration laboratories.

7.8 Formal Standardisation

Some communities have established strong community guidelines that are treated as de facto standards for observational processes, as discussed in the introductions to each of the main sections above. Other communities are considering formal standardisation, for example, there is a discussion within CEOS about whether the concepts of “analysis ready data” should be formalised in an ISO standard and they have already produced a specification requiring independent assessment called CEOS Analysis Ready Data for Land (CARD4L) and ESA has created its own ‘standards’ to assess and report usability of commercial satellite data products. In ocean observation, there are few guidelines focussing largely on long-established ECVs such as Sea Surface Temperature and Ocean Colour, but the community is recognising the value of such guidelines and is wanting to work towards establishing them for many other parameters and the metrology community should support these processes. The experience of metrology institutes with the CIPM MRA, and ISO 17025, can be valuable to these communities as such standards are introduced. Several of the people responding to our survey suggested that a role for metrology institutes is in supporting such standardisation and there was a request to include biological ECVs in such thinking, as well as the more obvious chemical and physical ones:

> "Yes, there is a gap in how biological ECVs are supported. There is a danger that biological ECVs will be developed independently rather than in association with the physical ECVs, creating redundancy and competition."

7.9 Link to Policy and Risk Management

The considerable investment in climate and ocean observation by global governments and international organisations is driven by a desire to understand our historic and current climate and to predict our future climate, so that this information may be used to inform policy and to manage risks. Section 2.1 described how the observations form part of a decision system and how there were two 'feedback loops'. The slower feedback loop (timescales of decades) is to monitor the Earth's environment as it responds to our anthropogenic forcing. The faster feedback loop (timescales of years) is to monitor anthropogenic changes: the (change in) GHGs emitted into the atmosphere and the land use changes that increase or decrease natural carbon sinks. This faster feedback loop is controlled through international treaties that come with requirements for regular reporting of national "inventories". Section 3.1.1 discussed how metrology can support the establishment of robust national inventories that can be meaningfully compared and combined globally and over time.

When observations (of both the climate system and of anthropogenic emissions and land use changes) are used to inform the policy of governments and of commercial companies, there is a translation from "measurement and model uncertainty" to "likelihood" and "risk". This translation has to be informed and traceable, and yet provided in a way that is meaningful to governments, economic markets and companies. The IPCC uses a defined vocabulary [116] to describe both uncertainty and confidence, using terms such as "very likely" (90 % - 100 % probability) and confidence (low, medium or high, based on level of disagreement and levels of evidence). But there is considerable criticism of these terms [117], suggesting that they are inconsistently applied and misunderstood by readers. In the BBC Podcast "How they made us doubt everything" [118], it was suggested that poor explanation of these terms by scientists, along with misuse by climate change deniers, contributed to a misunderstanding by the public of the reality of climate change.

Work is needed to consider how concepts of metrological uncertainty, as well as confidence in models and inference, can be expressed in ways that can be understood by a wide range of communities and interested people, including the general public. Metrologists must be part of such conversations.

More generally, metrology can add value to all parts of the chain described in Section 2.1 and shown in Figure 2. The obvious role of metrology is in the observations – both of the climate system (long loop) and of anthropogenic emissions and land use changes (short loop). But the metrological method and its focus on traceability – through uncertainty analysis and comparisons – can also support the assimilation of those measurements into models and the interpretation of those models into risk. This is a role for data scientists working within metrology institutes.

7.10 Summary of General Needs for Metrology

General Metrology Challenges for Climate and Ocean Observations
Performing metrological analysis on historical data and supporting data rescue. Investigating historical data – from networks, individual measurement records and early satellites to improve the calibration, perform comparisons, understand measurement methods and establish improved traceability and uncertainty evaluations based on modern knowledge. In this, supporting the development of robust uncertainty and covariance analysis for all observations used in reanalyses.
Metrological framework for model uncertainties. Establishing methodologies to evaluate uncertainties in ECVs processed partially or fully from models, in reanalyses and in climate models, including development of community-focused good practice guides.
Calibration methods and data processing techniques for low-cost sensor networks. Providing fit-for-purpose low-cost calibration to low-cost sensors and evaluating uncertainties and outliers in deployed low-cost sensor data.
Defining vocabularies and data-science ontologies for key metrological concepts. Working within communities to make metrological vocabulary more consistent, and ensuring formal ontologies contain enough terms to describe uncertainty robustly.
Providing training in metrological techniques. Theoretical training in concepts (e.g. uncertainties, traceability, comparison) and in practical training in calibration methods.
Supporting the development of QA frameworks and formal standardisation for networks. Working with community committees to establish metrologically-robust frameworks for network operation and data handling.
Support discussions on how uncertainty should be translated to commercial and societal risk. Create links from the observations through the models to societal and commercial decisions. Engage in the ongoing dialogue about how uncertainty should be treated in each stage of the value chain.

8 CONCLUSIONS AND FUTURE WORK

As this report has shown, metrology has a lot to offer climate and ocean observation. Metrology institutes have a valuable role for traditional metrological service provision – the calibration of instruments used in observation networks and satellite sensors and provision of reference materials, as well as traditional metrological research – development of new transfer standards, calibration procedures that match specialist instruments (for example, the broad field of view of a radiometric satellite sensor), and clarification of the definition of measurands (e.g. ocean salinity and pH), as well as supporting the standardisation of measurement processes and supporting the development of rigorous uncertainty analysis for observations. Metrologists are already working in strong collaboration with community experts in many of these areas, as described in many of the case studies above, and in our survey, there was an overwhelmingly positive response to the questions relating to existing and potential future collaboration with metrologists:

> "The interaction is very good, people have a high commitment, the NMI is highly supportive."

and:

> "Yes, I would welcome further engagement with the metrology community."

This report has shown several areas where further development and research is needed. There are many quantities that do not have in-field traceability to SI or even to a community agreed reference, further quantities where the measurand is not clearly defined, or the measurement procedure not standardised. Communities are asking for support from metrologists for all these challenges and for the development of rigorous uncertainty budgets, including information about the error correlation structures from observation to observation, to create the covariance matrices that data assimilation into models requires.

Metrology's role can, however, extend far beyond this. The metrological principle of "traceability", with its underpinning concepts of documentation and review, calibration chains, comparison and rigorous uncertainty analysis, can also be applied to the processing of raw measurements through complex processing chains to ECV quantities. In some areas metrologists, including data scientists working in metrological institutes, are already developing ways of applying metrological principles to observations that include modelling as part of the "retrieval", for example when correcting for atmospheric effects in satellite observations of land or ocean ECVs. There is a lot more that can be done.

Metrologists are also supporting the provision of retrospective traceability and new uncertainty budgets for historical observations, using comparisons of older instruments (or methods) with more recent ones and re-evaluating the measurement methodologies used in the 19th century and earlier, in in situ observations and in the 1980s and even 1970s for satellites. Here too, measurements and models are combined.

Metrology can also contribute to the broader discussions – and support the development of frameworks for how observational data are stored, for bringing uncertainty and traceability concepts into metadata and ontology standards, and to define the information required for long term data preservation. Metrology can bring the ideas of the MRA into observation communities – running community comparisons and establishing methods for peer review or auditing of the data that goes into ECVs. Metrology can support wider discussions about how "uncertainty" concepts are described to politicians, financial services and the general public.

In this report we have identified numerous 'urgent' needs which present varying degrees of challenge to the metrology community in terms of ability to achieve technically and obtaining the resources needed to address. In many cases they will require the development of new skills and creation of new partnerships. Prioritisation of actions and resources to address these needs in terms of achievability and criticality will be a challenge, as will coordination between the metrology community to maximise coverage and minimise unnecessary duplication of effort. Continued support and input from the climate and ocean user community to help with prioritisation as the EMN and its members try to develop a strategic research agenda to meet the needs will be welcomed and necessary. On-going actions of user communities to raise awareness and demand at governmental levels will also be needed to help increase the resources allocated to these tasks – and indeed inclusion of resources for metrology within community research programs to align with what NMI/DIs can bring will be essential.

Within this report we have talked of “the metrology community” as separate from those making and using climate observations. These are shorthand terms. We recognise that many metrologists have become sufficiently involved in observations such that they can be considered “climate scientists”, or oceanographers and scientists calibrating and operating sensors and sensor networks are metrologists. The observation of the Earth for climate science and oceanography is a multidisciplinary activity, and metrologists are one part of that broader community. We hope that those connections will strengthen.

At the time we finish this report, the Covid-19 crisis is still very present. There are lessons to be learnt on how policy, the public and science have interacted during that crisis. There have been politicians attempting evidence-based policies that “follow the science”, while the science is still unknown and regularly changing. There have been deniers who have exaggerated minor inconsistencies and challenged quality – for example the high level of criticism of the software quality for the (academic code) written by Imperial College London epidemiologists, or the misinterpretation by the press of normal scientific discussion between statisticians in Berlin. What is clear is that crisis management needs robust, high quality science backed by data and observations. The climate crisis is looming behind the Covid-19 crisis, and will be a bigger and more complex challenge to face in the decades to come. Humanity needs data and models backed by strong metrological principles.

The EMN for Climate and Ocean Observation will use this stakeholder needs review report as the basis for discussions to establish a strategic research agenda for the metrology institutes of Europe to respond to the challenges described here.

8.1 Summary of Identified Metrological Challenges

Metrology Challenges for Observations of Atmosphere ECVs
Fit-for-purpose working standards at appropriate concentrations, ensuring an unbroken SI-traceable calibration chain. There is also a need to improve the sampling and analytical methods for ambient measurements and to assess relevant influence parameters (typically aerosol and ozone precursors such as VOCs or NO_2)**.**
Certified reference materials for newly emitted halogenated compounds lacking standards (e.g. greenhouse gases such as HCFCs). Reference materials and a traceability chain for new measurements of isotopic composition and atmospheric tracers (e.g. dissemination of N_2O) must also be established**.**
Development of calibration procedures for aerosol properties using filter-based absorption photometers to provide a metrological framework for aerosol metrics beyond PM2.5 and PM10, and to generate source-specific reference aerosols in the laboratory.
Improve reference methods and instrumentation, typically for humidity measurements in the upper troposphere/lower stratosphere (less than 10 µmol/mol) under adapted environmental conditions (e.g. low pressure and temperature).
Support the establishment of the surface reference network by the GCOS (similar to GRUAN) as the top level of the WMO Integrated Global Observing System prescribed tiered approach in conjunction with the launch of the Global Basic Observing Network.
Metrological support for comparisons, particularly for challenging measurements lacking well-defined SI-traceability, and to compare surface, upper-air and satellite measurements (i.e. where there are very different traceability chains), including on site comparisons with metrological rigour for extreme environments and challenging locations e.g. cryosphere and high mountains.
Metrology support for specific initiatives focussed on the cryosphere, high mountains and urban areas. There is a need for metrologists to participate in multidisciplinary partnerships focussed on observations and predictions in key climate areas (e.g to participate in the establishment of an integrated high-mountain observation, prediction and services initiative).
Metrological support for field calibrations and measurements, including guidelines for using measuring devices on site including environmental influences and their uncertainty contribution.
Improved metrological characterisation of spectral parameters for chemical compounds (e.g. absorption cross-section, spectral line)**.**

Metrology Challenges for Observations of the Ocean
Definition of proper measurands and fit-for-purpose high order and working standards that ensure unbroken SI-traceable calibration chains. Currently, some of the ocean ECVs and EOVs are not defined in terms of SI units (e.g. pH, salinity). This makes it difficult to compare results obtained in different time and places, particularly when technology breaks occur.
Certified reference materials are essential tools to ensure the metrological traceability of results via the calibration of instruments, or to validate analytical measurement methods. Currently very few reference materials exist for some of the ocean ECVs and EOVs (e.g. inorganic carbon variables, pCO2, TA, pH) and most of them are not certified by NMIs/DIs.
Development of a metrologically based QA/QC framework and associated tools to facilitate field measurement reliability and consistent uncertainties. Currently, few oceanographic institutions are familiar with ISO 17025 accreditation. A scheme could be created on the example of QA4EO, establishing guidelines written in collaboration between the oceanography and metrology communities (see case study on QA4EO and its implementation in Section 6).
Organisation of interlaboratory comparisons for in situ measurements following metrological best practice to establish ‘degrees of equivalence’ and biases to enable international interoperability and harmonisation for long term comparability.
Fit-for-purpose uncertainties for in situ measurements, including training courses: GCOS requirements set stringent target uncertainties for many of the ECVs which are close to the level of primary standards. In contrast to this demand, assignment of uncertainties according to metrological concepts is not well established in oceanography.
Moving beyond best practice guidance documents and standard measurement procedures to international documentary standards, which can provide longer stability of measurement procedures over time.
On-board calibration for underwater instruments mounted on research vessels continuously measuring oceanographic parameters such as temperature, salinity, pressure, sound speed and bathymetry to ensure traceability and accuracy of measurements over instruments’ lifetimes and to account for environmental conditions and for their operation in dynamic mode.

Metrology Challenges for In Situ Observations of Land ECVs
Support to establish and assess traceability and associated uncertainty to community agreed (ideally SI) references for measuring systems under operational conditions.
Support to write documentary measurement procedures and best practice in a metrologically robust manner.
Mathematics to facilitate representativeness of observation from multiple samples/sites at both single locations and as part of a network.

Metrology Challenges for Remote Sensing
Pre-flight calibration standards and methods to enable SI-traceable uncertainties that commensurate with the needs of climate, available for cost/time efficient calibrations under operational conditions at industry/academic locations. These should cover the needs of all sensor domains e.g. passive optical through to active microwave, and address all necessary parameters and observational platforms e.g. space through to surface networks.
On-board calibration standards and methods to enable SI-traceable uncertainties that commensurate with the needs of climate to be achieved. This requires NMIs to support the transition of terrestrial techniques to orbit, assess degradation and uncertainty estimates for their use for all technology domains.
Establishing FRM quality test-sites/measurements for post-launch Cal/Val. This includes enabling SI-traceability in the field, uncertainty evaluation of in situ and its propagation/representativeness to the satellite all domains.
Development of a metrologically-based QA framework and associated methods/tools to facilitate evaluation and consistent reporting of end-to-end uncertainty of Level 1 and consequential higher-level data products. Including training support on uncertainty evaluation.
Metrological assessment of uncertainty of models & algorithms particularly those required to transform top-of-atmosphere measurands to bottom-of-atmosphere parameters, including support to developers, guidance on assessment, including the challenges of ‘machine learning’ methods. Means to establish uncertainty characterisation and representation for ‘classification’ systems e.g. land cover type, cloud masks etc. is also needed.
Comparisons and guidance for organisation of comparisons of community ‘measuring systems’. Comparison of satellite-to-satellite and satellite-to-ground systems following metrological best practice to establish ‘degrees of equivalence’ to enable international interoperability and harmonisation for long-time base FDRs.
Methods to establish metrologically robust FDRs and CDRs. How to combine both similar (e.g. sat series) and differing (different sensor designs) sources of data to create long-time base (multi-decadal) series with associated uncertainties. Facilitating interoperability (bias removal).

General Metrology Challenges for Climate and Ocean Observations
Performing metrological analysis on historical data and supporting data rescue. Investigating historical data – from networks, individual measurement records and early satellites to improve the calibration, perform comparisons, understand measurement methods and establish improved traceability and uncertainty evaluations based on modern knowledge. In this, supporting the development of robust uncertainty and covariance analysis for all observations used in reanalyses.
Metrological framework for model uncertainties. Establishing methodologies to evaluate uncertainties in ECVs processed partially or fully from models, in reanalyses and in climate models, including development of community-focused good practice guides.
Calibration methods and data processing techniques for low-cost sensor networks. Providing fit-for-purpose low-cost calibration to low-cost sensors and evaluating uncertainties and outliers in deployed low-cost sensor data.
Defining vocabularies and data-science ontologies for key metrological concepts. Working within the communities to make metrological vocabulary consistent, and ensuring formal ontologies contain enough terms to describe uncertainty robustly.
Providing training in metrological techniques. Theoretical training in concepts (e.g. uncertainties, traceability, comparison) and in practical training in calibration methods.
Supporting the development of QA frameworks and formal standardisation for networks. Working with community committees to establish metrologically-robust frameworks for network operation and data handling.
Support discussions on how uncertainty should be translated to commercial and societal risk. Create links from the observations through the models to societal and commercial decisions. Engage in the ongoing dialogue about how uncertainty should be treated in each stage of the value chain.

9 References

[1] “UN SDGs,.” www.un.org/sustainabledevelopment/.

[2] JCGM, “VIM: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition), JCGM 200:2012, Sevres, Paris; (2012),” 2008. https://www.bipm.org/en/publications/guides/#vim.

[3] “CIPM: Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes, Sevres, Paris (14 October 1999);” www.bipm.org.

[4] “GUM : JCGM 100:2008 Guide to the expression of Uncertainty in Measurement Sevres, Paris; (2012);” www.bipm.org.

[5] “AR5 Climate Change 2014: Mitigation of Climate Change — IPCC.” https://www.ipcc.ch/report/ar5/wg3/ (accessed Jun. 02, 2020).

[6] *United Nations / Framework Convention of Climate Change: Adoption of the Paris Agreement, 21 Conference of Parties, Paris.* 2015.

[7] “GCOS.” https://gcos.wmo.int/en/home.

[8] S. Bojinski, M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, “The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy,” *Bull. Am. Meteorol. Soc.*, vol. 95, no. 9, pp. 1431–1443, Sep. 2014, doi: 10.1175/bams-d-13-00047.1.

[9] “Special Report on the Ocean and Cryosphere in a Changing Climate.” https://www.ipcc.ch/srocc/ (accessed Sep. 11, 2020).

[10] OECD, “The Ocean Economy in 2030,” Apr. 27, 2016. https://www.oecd-ilibrary.org/economics/the-ocean-economy-in-2030_9789264251724-en.

[11] “Copernicus.” https://www.copernicus.eu/en.

[12] “WMO Global Atmosphere Watch (GAW) Implementation Plan: 2016-2023 | World Meteorological Organization.” https://public.wmo.int/en/resources/library/wmo-global-atmosphere-watch-gaw-implementation-plan-2016-2023 (accessed May 28, 2020).

[13] C. D. Keeling *et al.*, “Exchanges of atmospheric CO2 and 13CO2 with the terrestrial biosphere and oceans from 1978 to 2000.,” San Diego. doi: http://doi.org/10.6075/J08W3BHW.

[14] “Monitoring and verification of long term UK atmospheric measurement of greenhouse gas emissions - GOV.UK.” https://www.gov.uk/government/publications/uk-greenhouse-gas-emissions-monitoring-and-verification (accessed May 28, 2020).

[15] “2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories — IPCC.” https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/ (accessed May 28, 2020).

[16] A. Crotwell, A. Crotwell, M. Steinbacher, W. M. O. (WMO), and O. G. G. and R. M. T. (GGMT-2017) 19th WMO/IAEA Meeting on Carbon Dioxide, *19th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2017)*. Geneva: WMO, 2018.

[17] I. Levin, S. Hammer, E. Eichelmann, and F. R. Vogel, “Verification of greenhouse gas emission reductions: the prospect of atmospheric monitoring in polluted areas,” *Philos.*

Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 369, no. 1943, pp. 1906–1924, May 2011, doi: 10.1098/rsta.2010.0249.

[18] B. D. Hall, E. J. Dlugokencky, S. A. Montzka, G. Dutton, J. Mühle, and J. W. Elkins, "Long-lived greenhouse gases," in *State of the Climate in 2012*, 94th ed., J. Blunden and D. S. Arndt, Eds. Bull. Am. Meteorol. Soc, Boston, Mass., 2012, pp. S31–S33.

[19] M. Rigby *et al.*, "Recent and future trends in synthetic greenhouse gas radiative forcing," *Geophys. Res. Lett.*, vol. 41, no. 7, pp. 2623–2630, Apr. 2014, doi: 10.1002/2013GL059099.

[20] R. G. Prinn *et al.*, "History of chemically and radiatively important atmospheric gases from the Advanced Global Atmospheric Gases Experiment (AGAGE)," *Earth Syst. Sci. Data*, vol. 10, no. 2, pp. 985–1018, Jun. 2018, doi: 10.5194/essd-10-985-2018.

[21] M. K. Vollmer, S. Reimann, M. Hill, and D. Brunner, "First observations of the fourth generation synthetic halocarbons HFC-1234yf, HFC-1234ze(E), and HCFC-1233zd(E) in the atmosphere," *Environ. Sci. Technol.*, vol. 49, no. 5, pp. 2703–2708, Mar. 2015, doi: 10.1021/es505123x.

[22] *European Air Quality Directive 2008/50/EC*. 2008.

[23] *European Directive 2001/81/EC*. 2001.

[24] "ACTRIS station requirements." https://wmo-gaw-wcc-aerosol-physics.org/files/wp3.1-requirements-for-data-submission-and-reporting-reimann.pdf (accessed May 28, 2020).

[25] G. C. Rhoderick *et al.*, "Stability of gaseous volatile organic compounds contained in gas cylinders with different internal wall treatments," *Elem Sci Anth*, vol. 7, no. 1, p. 28, Jul. 2019, doi: 10.1525/elementa.366.

[26] J. Englert *et al.*, "Preparation and analysis of zero gases for the measurement of trace VOCs in air monitoring," doi: 10.5194/amt-2017-412.

[27] "MetNO2 Project." http://empir.npl.co.uk/metno2/.

[28] BIPM and WMO-GAWG, "BIPM-GAWG Particulate Workshop." http://www.bipm.org/ws/AllowedDocuments.jsp?ws=AEROSOLS .

[29] "European Union Air Quality Standards." https://ec.europa.eu/environment/air/quality/standards.htm.

[30] "AEROMET project." http://www.aerometproject.com/.

[31] M. Graf, L. Emmenegger, and B. Tuzson, "Compact, circular, and optically stable multipass cell for mobile laser absorption spectroscopy," *Opt. Lett.*, vol. 43, no. 11, p. 2434, Jun. 2018, doi: 10.1364/ol.43.002434.

[32] P. M. Hundt *et al.*, "Multi-species trace gas sensing with dual-wavelength QCLs," *Appl. Phys. B Lasers Opt.*, vol. 124, no. 6, pp. 1–9, Jun. 2018, doi: 10.1007/s00340-018-6977-y.

[33] J. W. Lovell-Smith *et al.*, "Metrological challenges for measurements of key climatological observables. Part 4: atmospheric relative humidity," *Metrologia*, vol. 53, no. 1, p. R40, Dec. 2016, doi: 10.1088/0026-1394/53/1/R40.

[34] W. M. O. (WMO) and W. M. O. (WMO), *Guide to Instruments and Methods of Observation*, 2018 Edition. Geneva: WMO, 2018.

[35] R. Feistel *et al.*, "Metrological challenges for measurements of key climatological

observables: oceanic salinity and pH, and atmospheric humidity. Part 1: overview," *Metrologia*, vol. 53, no. 1, p. R1, Dec. 2016, doi: 10.1088/0026-1394/53/1/R1.

[36] WMO-CIMO, "Report of the First Session of CIMO Expert Team on Standardization," 2012. [Online]. Available: http://www.wmo.int/pages/prog/www/IMOP/reports/2012/Final-Report_ET-Stand-1.pdf.

[37] WMO-CIMO, "Commission for Instruments and Methods of Observation Fifteenth session," 2010. Accessed: Sep. 16, 2020. [Online]. Available: https://www.wmo.int/pages/prog/www/CIMO/CIMO15-WMO1064/1064_en.pdf.

[38] "BIPM CCT Strategy Document." https://www.bipm.org/utils/en/pdf/CCT-strategy-document.pdf.

[39] C. Musacchio, G. Coppa, and A. Merlone, "An experimental method for evaluation of the snow albedo effect on near-surface air temperature measurements," *Meteorol. Appl.*, vol. 26, no. 1, pp. 161–170, Jan. 2019, doi: 10.1002/met.1756.

[40] WMO, "WMO High Mountains Summit Call to Action." https://highmountainsummit.wmo.int/en/call-action.

[41] WMO-CIMO, "WMO-CIMO: Report of the 16th Session," Geneva, 2014. [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=5576.

[42] P. W. Thorne *et al.*, "Toward an integrated set of surface meteorological observations for climate science and applications," *Bull. Am. Meteorol. Soc.*, vol. 98, no. 12, pp. 2689–2702, Dec. 2017, doi: 10.1175/BAMS-D-16-0165.1.

[43] E. J. M. Van Den Besselaar *et al.*, "International climate assessment & dataset: Climate services across borders," *Bulletin of the American Meteorological Society*, vol. 96, no. 1. American Meteorological Society, pp. 16–21, Jan. 01, 2015, doi: 10.1175/BAMS-D-13-00249.1.

[44] P. W. Thorne *et al.*, "Towards a global land surface climate fiducial reference measurements network," *Int. J. Climatol.*, vol. 38, no. 6, pp. 2760–2774, May 2018, doi: 10.1002/joc.5458.

[45] WMO, "Global Basic Observing Network (GBON)." https://www.wmo.int/pages/prog/www/wigos/GBON.html.

[46] WMO-CIMO, "CIMO: Report of the 17th Session," WMO, Geneva, 2018. Accessed: Sep. 28, 2020. [Online]. Available: https://library.wmo.int/index.php?lvl=notice_display&id=20696#.XUmIbPIzbIU .

[47] "GCOS-121 - Start of Implementation of GRUAN." https://www.gruan.org/documentation/gcos-wmo/gcos-121 (accessed Sep. 11, 2020).

[48] F. J. Immler, J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vömel, "Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products," *Atmos. Meas. Tech.*, vol. 3, no. 5, pp. 1217–1231, Sep. 2010, doi: 10.5194/amt-3-1217-2010.

[49] "GRUAN Homepage: About GRUAN." https://www.gruan.org/network/about-gruan (accessed Jun. 05, 2020).

[50] "GAIA-CLIM home page | Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring." http://www.gaia-clim.eu/ (accessed Jun. 05, 2020).

[51] M. van Weele, "Gaps Assessment and Impacts Document GAIA-CLIM Gap Analysis for

Integrated Atmospheric ECV Climate Monitoring,” 2015. doi: http://www.gaia-clim.eu/system/files/document/d6.11.pdf.

[52] “ATLANTOS.” http://www.atlantos-ocean.org/.

[53] “QARTOD: Quality assurance of real time ocean data.” https://ioos.noaa.gov/project/qartod/.

[54] “Ocean Best Practices.” https://www.oceanbestpractices.org/about/.

[55] J. Pearlman *et al.*, “Evolving and Sustaining Ocean Best Practices and Standards for the Next Decade,” *Front. Mar. Sci.*, vol. 6, no. JUN, p. 277, Jun. 2019, doi: 10.3389/fmars.2019.00277.

[56] GOOS, “GOOS Essential Ocean Variables.” https://goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114.

[57] “EOOS. EOOS Strategy 2018 – 2022.” www.eoos-ocean.eu.

[58] “EOOS Implementation Plan 2018 – 2022.” www.eoos-ocean.eu.

[59] “European Marine Sensors Calibration Network | JPI OCEANS.” http://www.jpi-oceans.eu/european-marine-sensors-calibration-network (accessed Aug. 06, 2020).

[60] A. G. Dickson *et al.*, “Metrological challenges for measurements of key climatological observables. Part 3: seawater pH,” *Metrologia*, vol. 53, no. 1, p. R26, Dec. 2016, doi: 10.1088/0026-1394/53/1/R26.

[61] R. Pawlowicz, R. Feistel, T. J. McDougall, P. Ridout, S. Seitz, and H. Wolf, “Metrological challenges for measurements of key climatological observables Part 2: oceanic salinity,” *Metrologia*, vol. 53, no. 1, p. R12, Dec. 2016, doi: 10.1088/0026-1394/53/1/R12.

[62] N. Fox, “Fiducial Reference Measurements for validation of Surface Temperature from Satellites (FRM4STS) D-160/D-170: FRM4STS Workshop Proceedings & FRM4STS Scientific Roadmap,” 2018.

[63] “What is Ocean Acidification?” https://www.pmel.noaa.gov/co2/story/What+is+Ocean+Acidification%3F (accessed Aug. 06, 2020).

[64] P. Tyack, “Global Ocean Observing System: EOV Specification Sheet: Ocean Sound.” https://goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=22567 (accessed Aug. 07, 2020).

[65] L. A. Levin *et al.*, “Global Observing Needs in the Deep Ocean,” *Front. Mar. Sci.*, vol. 6, p. 241, May 2019, doi: 10.3389/fmars.2019.00241.

[66] W. Wu, Z. Zhan, S. Peng, S. Ni, and J. Callies, “Seismic ocean thermometry,” *Science (80-.).*, vol. 369, no. 6509, pp. 1510–1515, Sep. 2020, doi: 10.1126/SCIENCE.ABB9519.

[67] EURAMET, “Infra-AUV project.” https://www.euramet.org/research-innovation/search-research-projects/details/project/metrology-for-low-frequency-sound-and-vibration/?tx_eurametctcp_project%5Baction%5D=show&tx_eurametctcp_project%5Bcontroller%5D=Project&cHash=51cf83cef6651516bac7ec6ec84b.

[68] K. G. Foote, S. P. Robinson, and P. D. Theobald, “Quantitative imaging of acoustic backscatter from the seafloor,” in *Proceedings of the Institute of Acoustics*, 2015, pp. 71–76.

[69] “Lurton, X.; Lamarche, G. (Eds) (2015) Backscatter measurements by seafloor-mapping

sonars. Guidelines and Recommendations. 200p." http://geohab.org/wp-content/uploads/2014/05/BSWGREPORT-MAY2015.pdf.

[70] K. G. Foote, "Standard-Target Calibration of Active Sonars Used to Measure Scattering: Principles and Illustrative Protocols," *IEEE J. Ocean. Eng.*, vol. 43, no. 3, pp. 749–763, Jul. 2018, doi: 10.1109/JOE.2017.2713538.

[71] I. Boyd *et al.*, "An International Quiet Ocean Experiment," *Oceanography*, vol. 24, no. 2, pp. 174–181, Jun. 2011, doi: 10.5670/oceanog.2011.37.

[72] "International Soil Moisture Network." https://ismn.geo.tuwien.ac.at/en/.

[73] "Quality Assurance for Soil Moisture Validation Service." https://qa4sm.eu/.

[74] C. Donlon, P. Goryl, P. Femenias, and S. Dransfeld, "Fiducial Reference Measurements (online presentation)." https://earth.esa.int/documents/700255/3425144/d2_10-20_donlon5eee.pdf (accessed Jun. 05, 2020).

[75] J. Mittaz, C. J. Merchant, and E. R. Woolliams, "Applying principles of metrology to historical Earth observations from satellites," *Metrologia*, vol. 56, no. 3, p. 032002, Jun. 2019, doi: 10.1088/1681-7575/ab1705.

[76] "CEOS CalValPortal." http://calvalportal.ceos.org/ (accessed Jun. 05, 2020).

[77] Z. Zhu *et al.*, "Benefits of the free and open Landsat data policy," *Remote Sens. Environ.*, vol. 224, pp. 382–385, Apr. 2019, doi: 10.1016/j.rse.2019.02.016.

[78] "QA4EO." www.qa4eo.org.

[79] "CCPR. Report CCPR-G2: Guidelines for CCPR key comparison report preparation." www.bipm.org/en/committees/cc/ccpr/publications-cc.html.

[80] J. Nightingale *et al.*, "Ten Priority Science Gaps in Assessing Climate Data Record Quality," *Remote Sens.*, vol. 11, no. 8, p. 986, Apr. 2019, doi: 10.3390/rs11080986.

[81] J. Nightingale *et al.*, "Quality Assurance Framework Development Based on Six New ECV Data Products to Enhance User Confidence for Climate Applications," *Remote Sens.*, vol. 10, no. 8, p. 1254, Aug. 2018, doi: 10.3390/rs10081254.

[82] "ESA FRM Programme," [Online]. Available: https://earth.esa.int/web/sppa/activities/frm.

[83] S. Sterckx *et al.*, "Towards a European Cal/Val service for earth observation," *International Journal of Remote Sensing*, vol. 41, no. 12. Taylor and Francis Ltd., pp. 4496–4511, Jun. 17, 2020, doi: 10.1080/01431161.2020.1718240.

[84] "GEOSS." www.earthobservations.org/geoss.php.

[85] "FIDUCEO." https://research.reading.ac.uk/fiduceo/.

[86] S. Sterckx *et al.*, "Towards a European Cal/Val service for earth observation," *Int. J. Remote Sens.*, vol. 41, no. 12, pp. 4496–4511, Jun. 2020, doi: 10.1080/01431161.2020.1718240.

[87] BIPM, "Rapport BIPM-2010/08: Report on the WMO-BIPM workshop on Measurement Challenges for Global Observation Systems for Climate Change Monitoring: Traceability, Stability and Uncertainty," Sevres, France, 2010. Accessed: Aug. 25, 2020. [Online]. Available: https://www.bipm.org/utils/common/pdf/rapportBIPM/RapportBIPM-2010-08.pdf.

[88] A. Barker *et al.*, "METROLOGY FOR CLIMATE Metrology priorities for the earth

observation and climate community," Teddington, UK, 2015. Accessed: Aug. 25, 2020. [Online]. Available: http://eprintspublications.npl.co.uk/6728/1/metrology_for_climate_report.pdf.

[89] X. Dong, "CEOS WGCV Microwave Sensors Subgroup Report (online presentation)." http://ceos.org/document_management/Working_Groups/WGCV/Meetings/WGCV-41/Presentations/CEOS-WGCV-41-SBR-03-MW.pdf (accessed Jun. 05, 2020).

[90] B. A. Wielicki *et al.*, "Achieving climate change absolute accuracy in orbit," *Bull. Am. Meteorol. Soc.*, vol. 94, no. 10, pp. 1519–1539, Oct. 2013, doi: 10.1175/BAMS-D-12-00149.1.

[91] N. Fox and P. Green, "Traceable Radiometry Underpinning Terrestrial- and Helio-Studies (TRUTHS): An element of a space-based climate and calibration observatory," *Remote Sens.*, vol. 12, no. 15, p. 2400, Aug. 2020, doi: 10.3390/RS12152400.

[92] M. Bouvet *et al.*, "RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range," *Remote Sens.*, vol. 11, no. 20, p. 2401, Oct. 2019, doi: 10.3390/rs11202401.

[93] N. Fox *et al.*, "Accurate radiometry from space: An essential tool for climate studies," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 369, no. 1953, 2011, doi: 10.1098/rsta.2011.0246.

[94] N. Fox *et al.*, "Traceable Radiometry Underpinning Terrestrial - and Helio- Studies (TRUTHS)," *Adv. Sp. Res.*, vol. 32, no. 11, pp. 2253–2261, Dec. 2003, doi: 10.1016/S0273-1177(03)90551-5.

[95] "Earth science and applications from space : national imperatives for the next decade and beyond /," National Academies Press, Washington, D.C., 2007. doi: https://lccn.loc.gov/2007936350.

[96] M. Dowell *et al.*, "Strategy Towards an Architecture for Climate Monitoring from Space," 2013. Accessed: Sep. 11, 2020. [Online]. Available: www.ceos.org;www.wmo.int/sat;http://www.cgms-info.org/.

[97] "TRUTHS - NPL." https://www.npl.co.uk/earth-observation/truths (accessed Jun. 05, 2020).

[98] A. Butz *et al.*, "Toward accurate CO2 and CH4 observations from GOSAT," *Geophys. Res. Lett.*, vol. 38, no. 14, Jul. 2011, doi: 10.1029/2011GL047888.

[99] H. Boesch, D. Baker, B. Connor, D. Crisp, and C. Miller, "Global Characterization of CO2 Column Retrievals from Shortwave-Infrared Satellite Observations of the Orbiting Carbon Observatory-2 Mission," *Remote Sens.*, vol. 3, no. 2, pp. 270–304, Feb. 2011, doi: 10.3390/rs3020270.

[100] Y. Govaerts, "Assessment of S3A/SLSTR Nadir and Oblique solar band calibration over Libya-4." https://www.eradiate.eu/resources/docs/reports/report-assessment_calibration_libya4-2.3-20191007.pdf (accessed Jun. 05, 2020).

[101] "Eradiate." https://www.eradiate.eu/site.

[102] G. Doxani *et al.*, "Atmospheric Correction Inter-Comparison Exercise," *Remote Sens.*, vol. 10, no. 3, p. 352, Feb. 2018, doi: 10.3390/rs10020352.

[103] D. R. Thompson *et al.*, "A unified approach to estimate land and water reflectances with uncertainties for coastal imaging spectroscopy," *Remote Sens. Environ.*, vol. 231, p.

111198, Sep. 2019, doi: 10.1016/j.rse.2019.05.017.

[104] C. D. Rodgers, “Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation,” *Reviews of Geophysics*, vol. 14, no. 4. John Wiley & Sons, Ltd, pp. 609–624, Nov. 01, 1976, doi: 10.1029/RG014i004p00609.

[105] J. Hobbs, A. Braverman, N. Cressie, R. Granat, and M. Gunson, “Simulation-based uncertainty quantification for estimating atmospheric CO2 from satellite data,” *SIAM-ASA J. Uncertain. Quantif.*, vol. 5, no. 1, pp. 956–985, Sep. 2017, doi: 10.1137/16M1060765.

[106] D. R. Thompson *et al.*, “Optimal estimation of spectral surface reflectance in challenging atmospheres,” *Remote Sens. Environ.*, vol. 232, p. 111258, Oct. 2019, doi: 10.1016/j.rse.2019.111258.

[107] C. J. Merchant *et al.*, “Uncertainty information in climate data records from Earth observation,” *Earth Syst. Sci. Data*, vol. 9, no. 2, pp. 511–527, Jul. 2017, doi: 10.5194/essd-9-511-2017.

[108] R. J. W. Brewin, L. de Mora, T. Jackson, T. G. Brewin, and J. Shutler, “On the Potential of Surfers to Monitor Environmental Indicators in the Coastal Zone,” *PLoS One*, vol. 10, no. 7, p. e0127706, Jul. 2015, doi: 10.1371/journal.pone.0127706.

[109] “A sky full of (commercial) eyes – Up to 1,300 Earth observation satellites in next 5 years – Space IT Bridge.” https://www.spaceitbridge.com/a-sky-full-of-commercial-eyes-up-to-1300-earth-observation-satellites-in-next-5-years.html (accessed Jul. 14, 2020).

[110] A. C. Lewis, E. Von Schneidemesser, and R. E. Peltier, “Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications.” Accessed: Jul. 14, 2020. [Online]. Available: https://www.ccacoalition.org/en/resources/low-cost-sensors-measurement-atmospheric-composition-overview-topic-and-future.

[111] W. M. O. (WMO) and W. M. O. (WMO), *Guidelines on Best Practices for Climate Data Rescue*, 2016 edition. Geneva: WMO, 2016.

[112] R. Böhm, P. D. Jones, J. Hiebl, D. Frank, M. Brunetti, and M. Maugeri, “The early instrumental warm-bias: A solution for long central European temperature series 1760-2007,” *Clim. Change*, vol. 101, no. 1, pp. 41–67, Sep. 2010, doi: 10.1007/s10584-009-9649-4.

[113] GCOS, “THE GLOBAL OBSERVING SYSTEM FOR CLIMATE: IMPLEMENTATION NEEDS GCOS-200 (GOOS-214).” Accessed: Sep. 04, 2020. [Online]. Available: https://gcos.wmo.int/en/gcos-implementation-plan.

[114] “NetCDF.” https://www.unidata.ucar.edu/software/netcdf/.

[115] “CF Standard Formal Names.” http://cfconventions.org/Data/cf-standard-names/56/build/cf-standard-name-table.html.

[116] J. Ridge *et al.*, “Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties Core Writing Team.” Accessed: Sep. 10, 2020. [Online]. Available: http://www.ipcc.ch.

[117] C. E. Adler and G. Hirsch Hadorn, “The IPCC and treatment of uncertainties: topics and sources of dissensus,” *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 5, no. 5, pp. 663–676, Sep. 2014, doi: 10.1002/wcc.297.

[118] BBC, “How they made us doubt everything.” https://www.bbc.co.uk/programmes/m000l7q1.

EURAMET e.V.
Bundesallee 100
38116 Braunschweig
Germany

Phone: +49 531 592 1960
Fax: +49 531 592 1969
E-mail: secretariat@euramet.org

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

EURAMET e.V. is a non-profit association under German law.