Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

Making better sense of the mosaic of environmental measurement networks: a system-of-systems approach and quantitative assessment

Peter W. Thorne[1], Fabio Madonna[2], Joerg Schulz[3], Tim Oakley[4,5], Bruce Ingleby[6], Marco Rosoldi[2], Emanuele Tramutola[2], Antti Arola[7], Matthias Buschmann[8], Anna C. Mikalsen[9], Richard Davy[9], Corinne Voces[1], Karin Kreher[10], Martine De Maziere[11], Gelsomina Pappalardo[2]

[1]Irish Climate Analysis Research UnitS, Department of Geography, Maynooth University, Maynooth, Ireland
[2]Consiglio Nazionale delle Ricerche, Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Potenza, Italy
[3]EUMETSAT, Darmstadt, Germany
[4]Global Climate Observing System Secretariat, Geneva, Switzerland
[5]Met Office, Fitzroy Road, Exeter, EX1 3PB, United Kingdom
[6]European Centre for Medium Range Weather Forecasting, Reading, United Kingdom
[7]Finnish Meteorological Institute
[8]Institute of Environmental Physics, University of Bremen
[9]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[10]BK Scientific, Mainz, Germany
[11] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

Correspondence to: Peter W Thorne (peter@peter-thorne.net)

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

Abstract. There are numerous networks and initiatives concerned with the non-satellite observing segment of Earth Observation. These are owned and operated by various entities and organisations often with different practices, norms, data policies etc. The Horizon 2020 project GAIA-CLIM is working to improve our collective ability to use an appropriate subset of these observations to rigorously characterise satellite observations. The first fundamental question is which observations from the mosaic of non-satellite observational capabilities are appropriate for such an application. This requires an assessment of the relevant, quantifiable aspects of the measurement series which are available. While fundamentally poor or incorrect measurements can be relatively easily identified, it is metrologically impossible to be sure that a measurement series is 'correct'. Certain assessable aspects of the measurement series can, however, build confidence in their scientific maturity and appropriateness for given applications. These are aspects such as that it is well documented, well understood, representative, updated, publicly available, maintains rich metadata etc. Entities such as the Global Climate Observing System have suggested a hierarchy of networks whereby different subsets of the observational capabilities are assigned to different layers based upon such assessable aspects. Herein, we make a first attempt to formalise both such a system-of-systems networks concept and a means by which to, as objectively as possible, assess where in this framework different networks may reside. In this study, we concentrate upon networks measuring primarily a sub-set of the atmospheric Essential Climate Variables of interest to GAIA-CLIM activities. We show assessment results from our application of the guidance and how we plan to use this in downstream exemplary applications of the GAIA-CLIM project. However, the approach laid out could be more widely applicable. If broadly adopted, the system-of-systems approach will have potential benefits in guiding users to the most appropriate set of observations for their needs, and in highlighting to network owners and operators areas for potential improvement.

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

1 Introduction

To observe aspects of the earth system in a variety of ways from satellite and non-satellite platforms is a necessary, but often insufficient, condition to enable full understanding. Observing in real-world conditions is a tough proposition. The real-world is not a lab where repeatable measurements under identical conditions are possible (e.g. Boumans, 2015). The real environmental state is constantly evolving both in space and in time. The real environmental exposure places practical limits on what can be measured, where it can be measured, and to what degree of fundamental quality it can be measured. It is therefore inevitable that there shall be a wide range of measurement systems and capabilities to meet distinct user needs and applications and that these may be further limited by a combination of technological, financial, geopolitical or logistical considerations. The challenge is how to make sense of such a mosaic of capabilities in order to properly inform data users of the most appropriate subset of measurements for their specific applications (e.g. Bojinski et al., 2014). Applications which, in many cases, were not foreseen when undertaking the original measurements. Here, we develop a system-of-systems framework approach to addressing this challenge and use as an illustrative case-study identification of suitable non-satellite atmospheric observational series which may be used to characterise satellite observations.

The Horizon 2020 project GAIA-CLIM aims to improve the usability of non-satellite measurements to characterise satellite measurements. A key step to achieving this is to identify, in as unambiguous manner as possible, those non-satellite measurements that are suitable for such an application. These reference observations must be sufficiently well characterised (JGCM, 2008, Immler et al., 2010, WMO/BIPM, 2010) that if a difference is found with the satellite data being compared, after accounting for inevitable co-location mismatch effects, we can be confident that the difference arises from the satellite and not from the comparator.

Were an appropriate, internationally accepted, method for identifying suitable non-satellite data sources for such a purpose available, then that would be used by the project. Surprisingly, to date there is no such accepted set of criteria by which to systematically evaluate the suitability for given applications across observing platforms and across networks using assessable metrics. Furthermore, while several international bodies refer to a system-of-systems observing architecture (e.g. GEO, 2016), there is still no formal definition of either the layers to such a system-of-systems or their defining and assessable characteristics. Different groups and domain-area experts have alighted on distinct conventions (GCOS, 2014), which makes it difficult for the potential users to quickly and easily assess which of the large set of potential measurements available are most appropriate to their application.

Therefore, we have attempted within GAIA-CLIM to put forth an initial definition of formal layers of a system-of-systems approach and a set of assessable metrics against which to categorise non-satellite observations, building upon existing efforts. In this paper, we outline the approach adopted by GAIA-CLIM and the results of its application to a restricted set of observations, covering a subset of the Global Climate Observing System (GCOS) atmospheric Essential Climate Variables (ECVs; Bojinski et al., 2014). Besides explicit reference to a system-of-systems approach in the peer-reviewed literature, by Seidel et al., 2009 and Bodeker et al., 2016, such a concept is also present in the grey literature, e.g. in several recent GCOS documents (GCOS, 2014, 2015) and a report by the US National Academy of Sciences (NAS, 2009). Although alluded to in these references, the defining characteristics of the layers are not clearly laid out. A formalised and systematically applied approach would help users and data providers to judge the usability of observational capabilities, and hence to use the right measurements for a specific application.

The paper summarises key results from two GAIA-CLIM deliverables (GAIA-CLIM, 2015, 2016; further details are available therein) and is structured as follows. Section 2 provides the proposed system-of-systems designation and describes the characteristics of each layer. Section 3 provides an overview on the assessment process adopted by GAIA-CLIM that builds upon the Climate Data Record (CDR)-based assessment proposed by CORE-CLIMAX (Schulz et al., in prep; which in turn build upon Bates and Privette (2012)), modified to account for distinctions between measurements and derived data products.

Section 4 describes initial results of an application to a selected set of networks suitable to meeting GAIA-CLIM's needs. Section 5 discusses caveats and potential future developments and Section 6 provides a summary.

2 System-of-systems approach adopted by GAIA-CLIM

The system-of-systems approach adopted is illustrated in Figure 1. This recognises that different observations exist for distinct purposes and the inevitable trade-off that exists between spatio-temporal observational density and aspects of observing quality. The designation of any candidate measurement series to a given layer should be a function of demonstrable measurement qualities such as: traceability, metadata, comparability, data completeness, documentation, record longevity, measurement program stability and sustainability, etc. (Section 3). Before this, however, the defining characteristics of each layer in Figure 1 need to be formally defined. These definitions build and expand upon GCOS (2014), with further details given in GAIA-CLIM (2015).

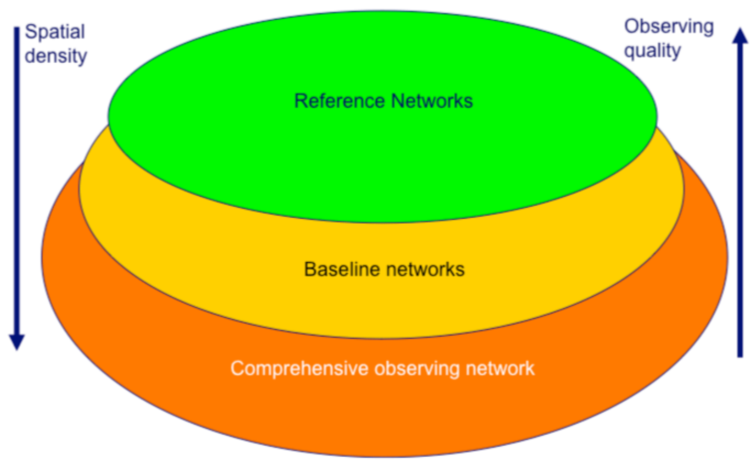

Figure 1. Proposed layers in a system-of-systems approach to be adopted within GAIA-CLIM arising from Seidel et al. 2009.

2.1 Reference observing networks

Reference observing networks provide metrologically traceable observations, with quantified uncertainty, at a limited number of locations, and/or for a limited number of observing platforms, for which traceability has been attained.

- The measurements are traceable through an unbroken processing chain (in which the uncertainty arising in each step has been rigorously quantified) to SI units, Common Reference Points defined by BIPM, or community recognised standards (ideally recognised by National Measurement Institutes), using best practices documented in the accessible literature (Immler et al., 2010). Dirksen et al. (2014) provides an example of the steps required to deliver such a product.
- Uncertainties arising from each step in the processing chain are fully quantified and included in the resulting data. Uncertainties are reported for each data point. Individual components of the uncertainty budget are available. Where uncertainties are correlated, these are appropriately handled.
- Full metadata concerning the measurements is captured and retained, along with the original raw data, to allow subsequent reprocessing of entire data streams as necessary.
- The measurement and its uncertainty are verified through complementary, redundant, observations of the same measurand on a routine basis.
- The observations program is actively managed and has a commitment to long-term operation.
- Change management is robust including a sufficient program of parallel and/or redundant measurements to fully understand any changes that do occur. Unnecessary changes are minimised.
- Measurement technology innovation is pursued. New measurement capabilities through new measurement techniques, or innovations to existing techniques, which demonstrably improve the ability to characterize the measurand, are encouraged. These innovations must be managed in such a way as to understand their impacts on the measurement series before they are deployed.

2.2 Baseline observing networks

Baseline observing networks provide long-term records that are capable of characterising regional, hemispheric and global-scale features. They lack the absolute traceability of reference observations.

- The baseline network is a globally or regionally representative set of observations capable of capturing, at a minimum relevant large-scale changes and variability. As such, a baseline network may be considered a minimum and highest priority subset of the comprehensive networks (Section 2.3), and should be actively curated and retained.
- The measurements are periodically assessed, either against other instruments measuring the same geophysical parameters at the same site, through comparisons to NWP / reanalyses, or, through inter-comparison campaigns to provide understanding of the relative performance of the different techniques in use. Ideally, such inter-comparisons should include reference-quality measurements.
- Representative uncertainties, that are based upon understanding of instrument performance or peer-reviewed literature are available.
- Metadata about changes in observing practices and instrumentation are retained.
- The observations have a long-term commitment.
- Changes to the measurement program are minimized and managed (by overlapping measurements, or measurements with complementary instruments over the change), with efforts made to quantify the effects of changes in an appropriate manner.

2.3 Comprehensive observing networks

Comprehensive observing networks provide high spatio-temporal density data information necessary for characterising local and regional features.

- The comprehensive networks provide observations at the detailed space and time scales required to fully describe the nature, variability, and change of a specific climate variable, if analysed appropriately. They include regional and national operational observing networks.

- Representative uncertainties based upon, e.g., instrument manufacturer specification and knowledge of operations should be provided. In their absence, gross uncertainties based upon expert or operator judgement should be provided.
- Metadata should be retained.
- Although encouraged, long-term operation is not required.

3. Objectively assessing measurement capabilities

The measurement system maturity matrix (MSMM) used herein, like its counterpart for CDRs developed under CORE-CLIMAX (Schulz et al., in prep), is a tool to assess various assessable facets of a measurement series or measurement network. It assesses to what extent measurement best practices have been met. The assessment can be performed either on individual instruments/sites or for entire networks. It should be stressed that a given measurement's maturity is distinct from its applicability to a given problem, where additional concerns, such as measurement location, frequency, scheduling etc. pertain. For example, a user interested in tropical processes cannot make use of a measurement in the polar regions and vice-versa. Such aspects are user specific, and cannot be captured within the matrices detailed herein.

3.1 Maturity assessment overview

There are six mandatory and one optional major categories where assessments are made, which overlap with, but are not identical to, those used to assess CDRs under CORE-CLIMAX (Schulz et al., in prep). Where the categories overlap, in many cases the guidance differs substantially to reflect the distinction between the measurements and derived CDRs. The assessment categories are:

- Metadata
- Documentation
- Uncertainty characterisation
- Public access, feedback, and update
- Usage

- Sustainability
- Software (optional, completed only where appropriate)

Within each category are a number of sub-categories. For each of these sub-categories, the assessment will assign a score from 1 to 6 (sometimes 6 is not used and/or 1 and 2 are identical criteria), reflecting the maturity of that aspect of the measurement system.

3.1.1 Maturity scores and the system-of-systems networks concept

The maturity can be considered in three broad categories that give information on the scientific grade and sustainability of the measurements being assessed.

- Maturity scores 1 and 2 establish **Comprehensive Measurement Capability** (CMC, Comprehensive network type measurements): The instruments are placed in the field and recording data, but may not be well curated or metrologically understood and calibrated.
- Maturity scores 3 and 4 establish a **Baseline Measurement Capability** (BMC, Baseline network type measurements): These measurements are better characterised and understood, and intended to be run for the long-term. However, they lack strict traceability and comparability.
- Maturity scores 5 and 6 establish a **Reference Measurement Capability** (RMC, Reference network type measurements): These measurements are very well characterised, with strict traceability and comparability, and robustly quantified uncertainties. The measurements are actively managed and curated, and envisaged as a sustained contribution.

3.1.2 Interpreting the maturity assessment results

Assessment of results may require expert interpretation for each assessed measurement series, because the circumstances under which the measurements were taken may affect what maturity level can be reasonably expected to be attained. All relevant sub-category scores should be considered. From the data provider's perspective, such an assessment may inform strategic developments and improvements to the measurement program. From the perspective of data users, the assessment should provide an indication of applicability to their intended use.

When considering an assessment of a network, in certain categories or sub-categories, it is likely to be appropriate to perform the exercise on a per-asset (instrument or site) basis, rather than a network-wide basis. This is particularly the case for the "Sustainability" category, but may also be applicable elsewhere if there are intra-network heterogeneities in protocols pertaining to, e.g., metadata, uncertainty quantification or documentation. In such cases, and where practical, the assessment should be performed individually on each unique subset and stored in the assessment report metadata. Both the network-wide mean score (or a representative score of 'core' sites) and the range of scores should then be reported in the summary. Such a refined assessment helps ensure both appropriate network sub-selection for certain applications, and a fair assessment, that may help network operators and coordinators identify and address network internal issues.

3.2 How to perform an assessment

An assessment should be conducted by an assigned leader, who organises the exercise, provides guidance to the participants, and collects and analyses the results. Where a substantive assessment of the state of multiple networks, instruments or sites is being organised, it is recommended to create an additional supplement of specific assessment criteria details or '*rules of the round*'. This guidance should be retained alongside the completed assessments, to permit full interpretation of the results.

3.3 Overview of assessment strands

Full guidance for assessment strands is given in GAIA-CLIM (2015), including detailed guidance notes to aid assessors. We reproduce here (lightly edited for clarity) the guidance given for the first sub-category of metadata standards in full; but thereafter provide solely a high-level overview of each category for brevity. All remaining sub-categories contain similar tables and assessment guidance to that shown in Section 3.3.1.1. Readers wishing to perform an assessment should refer to the full in-depth guidance (GAIA-CLIM, 2015) and any subsequent update thereto.

3.3.1 Metadata category

Metadata is 'data' about data, which should be standardised, as complete as possible, and adequately documents how the measurement series was attained. This involves aspects such as instrumentation, siting, observing practices, etc. The measurement system should use appropriate high-quality metadata standards, which permit inter-operability of metadata. If an International Organization for Standardization (ISO) standard is defined, then the assessment in future would be against such a standard. However, at the present time no such universally agreed standard exists that pertains across all aspects of EO science. There are emerging efforts under WIGOS (WIGOS, 2015a,b) to create universal metadata standards, and there are several *de facto* working standards such as CF-compliant file headers and formats. Unless and until an ISO standard is developed and applied, the assessors' judgement will be required as to the appropriateness of the standards being adhered to.

3.3.1.1 Example full assessment guidance – Metadata Standards

Standards – It is considered to be good practice to follow recognized metadata standards. These may differ depending upon the instrument or measurement program under consideration, and may be determined on a network / infrastructure-wide basis. As discussed previously, currently no ISO-standard for metadata exists. Table 1 provides high-level assessment guidelines associated with each score.

1	No standard considered
2	No standard considered
3	Metadata standards identified and/or defined and partially but not yet systematically applied
4	Score 3 + standards systematically applied at file level and collection level by data provider. Meets international standards
5	Score 4 + meta data standard compliance systematically checked by the data provider

6	Score 5 + extended metadata that could be useful but is not considered mandatory is also retained.

Table 1: The 6 maturity scores in metadata sub-category Standards.

Note that it is likely that this sub-category can only be fully assessed by the measurement initiator. An external assessment can be made by asking the data provider directly, or if the metadata and data are freely available from a portal (which would tend to indicate a mature measurement system). However, signs for used standards can be found by looking at the data record documentation and/or at a sample data file.

The assessment can be made as follows:
Score 1 and 2 (no distinction made in this case between these two levels): No standard is considered. Data are made available solely as are with at most the geographical measurement location, time of observation and instrument-type metadata provided that enables use, but prohibits measurement understanding.

Score 3: "Standard identified/defined" means that the measurement originator has identified or defined the standard to be used, but has not yet systematically applied it. The information about this can most often be found in format description documents, or from statements on web pages.

Score 4: A systematic application requires that you can find the relevant metadata protocol identifier and details in every file of the measurement product and descriptions.

Score 5: The measurement provider has implemented procedures to check the metadata contents. This could be ascertained by a check on consistency of metadata header information in individual data files.

Score 6: This score will be attained if, in addition to mandatory metadata, additional optional metadata is collected, retained and transmitted. This score may not apply to some data streams where all metadata is considered mandatory, but may help differentiate truly well performing measurement series in other cases, where metadata is differentiated into mandatory and optional classes, such as the WIGOS metadata standards (WIGOS, 2105a,b) for example.

3.3.1.2 Additional sub-categories for metadata

Collection Level metadata – these are attributes that apply across the whole of a measurement series, such as processing methods (e.g., same algorithm versions), general space and time extents, creator and custodian, references, processing history etc. Discovery metadata through e.g. use of digital object identifiers, can form part of this and ensure long-term discoverability. Collection level metadata allows others to find out what the measurement series contains, where it was collected, where and how the series is provided, and what usage restrictions apply.

File level attributes are those specific to the granularity of the data (on a per measurement basis) and vary with each measurement entity. The file level metadata includes such elements as time of observation, location, measurement units, measurement specific metadata, such as ground check data, measurement batch number, ambient conditions at time of observation, etc. Such metadata are necessary to understand and properly use the individual measurements.

3.3.2 Documentation

Documentation is essential for the effective use and understanding of a measurement record. Although the category has three sub-categories, it is possible that two or more of these may be covered by a single document.

Formal description of measurement methodology refers to a description of the physical and methodological basis of the measurements, network status (if applicable), processing of the raw data and dissemination. It shall often be used as a manual by the site technicians for how to take the measurements. This can cover such aspects as descriptions of measurement principles, methods of observation, calibration procedures, data filtering, data processing, corrections, aggregation procedures, data distribution etc.. Where software is involved in the processing of the data, its availability should be assured.

A **Formal validation report** contains details on the validation activities that have been undertaken to assess the fidelity/reliability of the measurement record. It describes uncertainty characteristics of the measurement record found through the application of uncertainty analysis (Section 3.3.3), and provides all relevant references.

Formal measurement series user guidance contains details necessary for users to discover and use the data in an appropriate manner. It includes aspects such as the technical definition of the measurement series, overview of instrumentation and methods, general quality remarks, validation methods and estimated uncertainty in the data, strengths and weaknesses of the data, format and content description, references, and processing details.

3.3.3 Uncertainty characterisation

This category assesses the practises used to characterise and represent uncertainty in a measurement series. Note that uncertainty nomenclature and practices must follow established definitions (JGCM, 2008) to attain a score of 5 or 6 in any of the sub-categories.

Traceability is attained if the measurement series can be related to stated references, through an unbroken chain of comparisons, and that these processing procedures all have stated / quantified uncertainties. To support a claim of traceability, the provider of a measurement must document the measurement process or system used to establish the claim, and provide a description of the chain of comparisons that were used to establish a connection to a particular stated reference. Full traceability on a sustained basis requires in-depth instrument understanding and regular comparisons to standards, and will typically involve and be certified by National Measurement Institutes.

Comparability considers the extent to which the measurement has been validated to provide realistic uncertainty estimates and stable operations through in-the-field comparisons. Such validation is substantively distinct from traceability in that it relates to a sustained program of comparison both in the measured environment and using lab-based experiments to ascertain potential biases, drifts and artefacts between two measurements. Unlike for traceability, the comparison need not be to a measure that itself is traceable. However, for the highest quality measurements such comparisons should be against measurements that are themselves traceable.

Uncertainty quantification evaluates the extent to which uncertainties have been fully quantified and their accessibility to users.

Routine quality monitoring is the monitoring of data quality while processing the data. Quality monitoring is a robust and quantitative measure of how closely an individual measurement conforms to an expectation against which the observations can be compared and assessed. Such quality monitoring helps to assess, in near real time, major issues with the measurements, and permits proactive management. It may lead to a stop and restart of processing activities or measurement series if any type of error is detected.

3.3.4 Public access, feedback and update

This category relates to archiving and accessibility of the measurement record, how feedback from user communities are established, and whether this feedback is used to update the measurement record. It also concerns version control, archival and retrieval of present and previous versions.

Access evaluates the ease of distributing the raw and processed data, documentation, and any necessary source code used to process the data from the raw measurement to the geophysical or radiance parameter space. Public access means that the data are available without restrictions for at least academic use, but such access may still be subject to a reasonable fee. The raw data may only be provided upon request, but a mechanism for requesting should be readily apparent in such cases. The highest scores in this category can only be attained for data provided free of charge without restrictions on use and re-use. Data provider here means either the data collector or organisations such as space agencies, national meteorological centres or research institutes.

User feedback is important for developers and providers of measurement records to improve quality, accessibility, etc. of a given measurement series. This category evaluates whether mechanisms are established to receive, analyse, and use such feedback. Feedback can reach a measurement provider in many ways, but needs to be organised in such a way that it can be used to improve a measurement record and/or the service around it.

Updates to record evaluates if data records are systematically updated when new observations or insights become available, or if this is done in *ad hoc* fashion if at all. More mature measurement series will tend to be updated in an operational manner that assures both their sustainability and their suitability for applications requiring reliable data updates.

Version control allows a user to trace back the different versions of algorithms, software, format, input and ancillary data, and documentation used to generate the measurement record under consideration. It allows clear statements about when and why changes have been introduced, and allows users to document the precise version of the data they used, thus enabling replication of users' analyses.

Long-term data preservation relates to the preservation of measurement series records. According to Long Term Data Preservation guidelines (http://earth.esa.int/gscb/ltdp/), an archive should keep more than one copy, use different media/technologies, and different locations. Most important is to retain the raw data and necessary metadata, which may allow subsequent reprocessing.

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

3.3.5 Usage

This category is related to the usage of measurement series in research applications and for decision support systems. Mature measurement series will have broad adoption and widespread and varied usage.

Research applications of a measurement series can be evaluated by its appearance in publications and citations of such publications.

Public and Commercial Exploitation covers any direct use in real-time monitoring, forecasts, infrastructure planning, support to agencies or other business areas such as insurance and indirect support, e.g., through citations in IPCC reports, to decision and policy making in socio-political contexts.

3.3.6 Sustainability

This category pertains to aspects of sustainability, and hence suitability, of any given measurement program for scientific, operational, and societal applications. For a measurement program to be used in critical applications, its long-term sustainability must be assured. Where an international measurement network is being assessed, the network shall typically consist of individual measurement sites operated by distinct legal entities, with distinct funding mechanisms, and in a variety of siting environments. In such cases, there are two options. One is to provide a typical score, that is representative of the network as a whole, but this is then not indicative of the maturity of individual contributing sites. The alternative, preferred option, is that this assessment be performed for each site, with the site-by-site scores retained as metadata associated with the assessment, and the range of scores highlighted appropriately in the assessment summary by providing both a mean value and the range.

Siting environment only applies to fixed measurement assets, for which observations are taken repeatedly from a single location (including weather balloons which originate from a constant location but may drift), or mobile observations using repeating transects. Non-repeating measurements made from aircraft and other mobile platforms should leave this entry blank, and use solely the remaining strands to assign a score under sustainability. Within this category, consideration is limited to the representativeness of the site / transect of its immediate surrounding environment / landscape.

Scientific and expert support evaluates the degree of scientific, technical and measurement science expertise that underpins the measurement programme.

Programmatic support assesses the long-term programmatic support that underpins the measurement program.

3.3.7 Software readiness (optional)

This major strand is optional, and shall apply only to those measurements where routine automated and substantive processing occurs from the raw measured data to the provided geophysical parameters of the measurement series. Cases where this would be appropriate would include measurement series where the directly measured parameter is a digital count, a radiance, a photon count or some other indirect proxy for the reported measurand, and processing exists to convert from the measured quantity to the reported quantity. Conversely, where the measurement constitutes a direct proxy for the measurand, such as a platinum resistance thermometer or anemometer, and the conversion is facile, the software readiness category is not appropriate.

Coding standards are a set of conventions/rules specific for a coding language, which describe style, practices and methods that greatly reduce the probability of introducing bugs. This is especially important in a team environment, or group collaboration, so that uniform coding standards are used, and helps to reduce oversight errors and save time for code reviews. It is key to assuring the maintainability of the code at reasonable cost. There are ISO standards available for software coding which may be applicable. If such ISO standards are to be used it should be agreed in the 'rules of the round'.

Software Documentation is key to ensuring usability, portability and operator understanding. This sub-category is concerned primarily with whether the code is documented with proper headers, change history, and sufficiently complete and understandable comments describing the processes.

Portability and numerical reproducibility concerns the usability of the software in different environments (different computing platforms such as Linux, Solaris, Mac OS, Windows etc. and different compilers such Intel, IBM, GNU, Portland, etc), and whether the results are numerically reproducible. It is important for migrating software from old to new computer systems and from one place to another.

Security is associated with software contents that either have the potential to destroy files and complete environments or are related to file transfer between compute environments. Both should not be contained in software. The security category also checks whether the file system can be accessed from outside, as this may hamper the integrity of the measurement series generation environment.

4. Results of the assessment performed by GAIA-CLIM participants

We identified a total of 54 plausible networks and two aircraft permanent infrastructures for EO characterisation in the context of GAIA-CLIM activities (Appendix A provides a full accounting of these). These networks are those based upon expert solicitation to be most likely to constitute baseline or reference-quality measurement systems according to the criteria put forth in Sections 2 and 3, and hence be usable in downstream applications within the project. The assessment results will thus *a priori* be likely to be at baseline or reference level relative to a holistic assessment of the entirety of the non-satellite observational capabilities. Such an assessment, while highly desirable, would constitute a far more substantive effort than was possible under GAIA-CLIM. We were able to complete, or solicit assessments by third party contributors to, 43 of these networks, or sub-networks as part of the same research infrastructure (Appendix A). The assessed networks cover a broad range of geographical locations (Figure 2). As expected, the most sparsely populated and remote regions of the Earth are where the density of measurement stations is lowest.

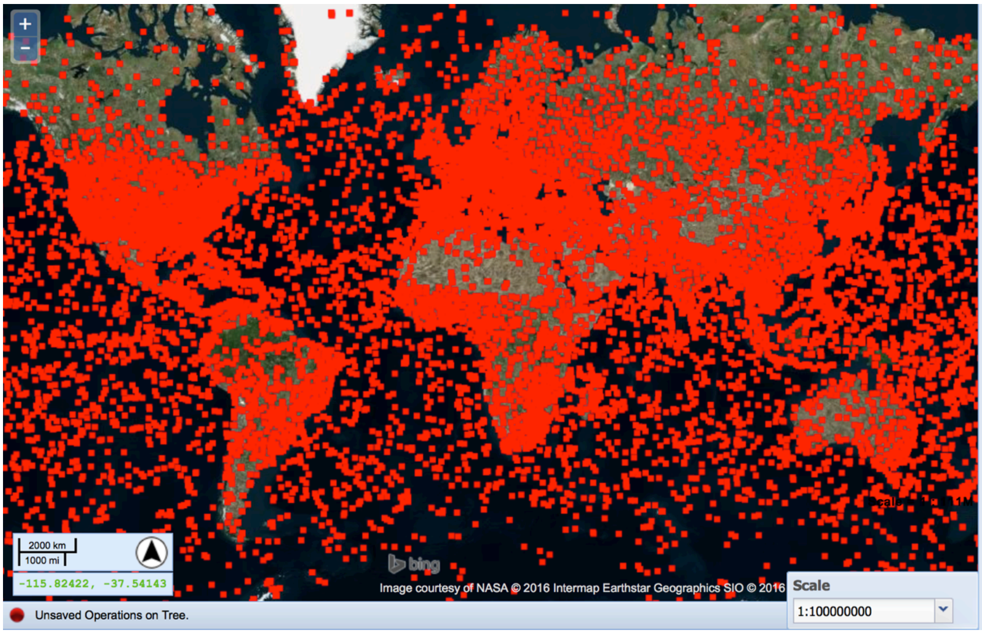

Figure 2. Sites of networks assessed for GAIA-CLIM. Each location denotes an observational asset capable of measuring one or more target ECVs at the surface, near surface or through much of the atmospheric column.

4.1 How the assessment was performed

Per the developed guidance (Section 3; GAIA-CLIM, 2015) a set of 'rules of the round' were agreed at the outset. The assessment was performed on a network-by-network basis given the available time and the project resource constraints. The maturity matrix collection has been carried out by co-authors based upon their individual areas of expertise and their involvement in several international measurement programmes and networks. Significant effort has been made to fill in the matrices consistently across networks. In those cases where filling in the matrix was considered challenging by co-authors, an assessment aided by the assessed network PI or other core members has been solicited (see acknowledgements). In such cases, the authors worked to fully support the network PIs to ensure a consistent compiling.

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access

An example of maturity matrix collection is provided in Figure 3 for the Network for the Detection of Atmospheric Composition Change (NDACC; NDACC, 2015) as filled in by NDACC working group co-chairs guided by BIRA-IASB colleagues, who are active participants in managing this network. The scores reported in Figure 3 show that NDACC can be considered, according to the MSMM, as a reference network in sub-categories like data traceability, but for the uncertainty quantification, the network is currently assessed as baseline level. All the matrices in the same form as shown in Figure 3 for NDACC are available on the GAIA-CLIM website (http://www.gaia-clim.eu/page/maturity-matrix-assessment).

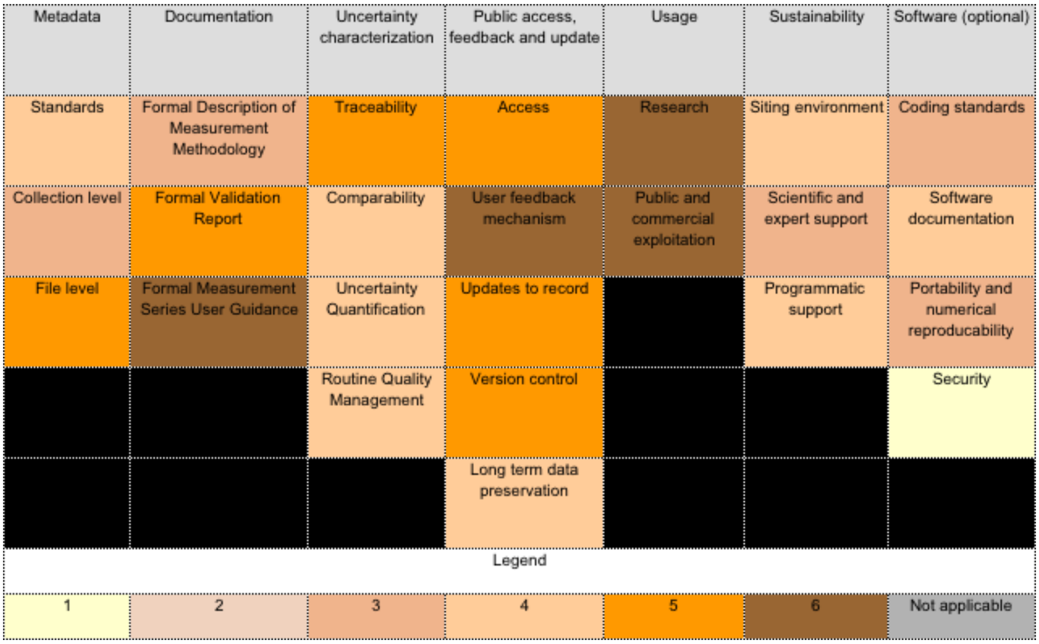

Figure 3. Example maturity matrix assessing the NDACC (Network for the Detection of Atmospheric Composition Change).

A summary of all results attained is given in Figure 4 (c.f. Figure 3). Here, the single collated and agreed set of assessed maturity scores is given per network, even if several contributions were solicited (see Section 4.2). Most networks completed assessments for at least all the mandatory assessment criteria. Usage and software, given the agreed rules of the round, have the lowest level of completion. A synthesis

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

NETWORK	METADATA			DOCUMENTATION			UNCERTAINTY CHARACTERISATION				PUBLIC ACCESS, FEEDBACK, AND UPDATE					USAGE		SUSTAINABILITY			SOFTWARE READINESS			
	Standards	Collection level metadata (including change records)	File level	Formal description of measurement methodology	Formal Validation Report	Formal Measurement series User Guidance	Traceability	Comparability	Uncertainty quantification	Routine Quality Monitoring	Public Access/Archive	User Feedback Mechanism	Updates to Record	Version control	Long-term data preservation	Research	Public and commercial exploitation	Siting environment	Scientific / expert support	Programmatic support	Coding standards	Software Documentation	Portability and Numerical Reproducibility	Security
ACTRIS surface	5	5	5	3	4	4	4	4	5	4	4	4	4	3	5	5	4	5	5	6	3	2	1	1
ACTRIS profiles	5	5	5	5	4	4	5	4	5	4	6	4	5	3	5	3	4	5	5	6	5	2	1	1
ADNET	6	6	5	5	6	5	5	4	5	5	6	6	5	6	6	5	4	5	5	6	6	6	6	5
AEROCAN	4	3	5	6	5	5	5	5	3	6	5	5	6	5	5	5	5	6	5	5				
AERONET_PHOTONS	4	3	5	6	4	5	5	5	3	6	5	5	6	5	5	5	5	6	5	5				
AGSNET_CSIRO	4	3	5	6	4	5	5	5	3	6	5	5	6	5	5	5	5	6	5	5				
AMeDAS	5	6	5	6	6	5	6	5	5	5	4	5	6	6	5	6	6	4	5	6	6	5	4	5
ARGO	6	6	5	6	4	4	4	3	2	6	6	4	6			5	3		5	6	6			
BSRN	6	6	5	5	5	5	6	6	6	6	6		6			5	3	6	5	6		6		
EARLINET	6	6	5	5	6	5	5	4	5	5	6	6	5	6	6	5	4	5	5	6	6	6	6	5
EMEP	4	6	5	4	6	6	5	4	5	5	4	2	4	4	5	6	5	6	5	6	3	5	5	2
EPA	3	4	5	5	2	3	4	3	3	5	6	2	5		5	2	4	6	3	5				4
EUREF	5	6	5	5	5	6	5	6	5	6	6	6	6	5	5	6	6	6	5	6	1	1	1	1
EUROSKYRAD	3	4	3	3	4	5	5	4	3	5	5	2	3	3	3	4	2	5	5	4		6	6	4
GAWPFR	4	2	5	5	5	4	4	5	2	5	4	2	4		5	5	2	4	5	5				
GPSMET	6	6	4	6	4	4	4	3	3	5	6	3	6	3	3	5	3	6	4	5	6	6	6	5
GRUAN	5	6	5	5	6	6	4	6	4	6	4	6	6	6	5	4	4	4	5		3			
GSN	6	6	5	3	3	3	3	3	3	5	6	3	6	2	2	4	3	3	4	5	6	5		
GUAN	6	6	5	5	3	3	3	3	3	5	6	3	6	2	2	4	3	5	4	5	6	5	3	4
ICOS (including InGOS)	6	3	5	4	5	2	5	5	4	4	1	2	1	1	5	5		5	5	5				
IGS	5	6	5	5	5	6	5	6	5	6	6	6	6	5	5	6	6	6	5	6	1	1	1	1
IMPROVE	3	4	5	5	2	3	5	3	4	6	6	3	5		5	2	4	6	4	5				5
MESONET	3	4	4	1	1	2	1	2	1	2	1	2				2	3	6	2	4				
MPLnet	5	4	5	6	5	5	4	3	6	5	6	5	2	6	6	6	3	6	5	5	2	3	3	5
MWRnet	4	4	5	4	5	5	4	3	4	4	4	4	4	3	4	5	5	6	5	6	4	4	4	2
NDACC	4	3	5	3	5	6	5	4	4	4	5	6	5	5	4	6	6	4	3	4	3	4	3	1
NOAA ESRL WV profiles	5	3	5	5	6	6	3	4	2	4	5	4	5	4	5	5		5	5	4				
NOAA ESRL O3 profiles	5	3	5	3	4	3	3	3	1	4	4	4	5	3	5	5		5	5	4				
NOAA ESRL O3 dobson	5	3	2	5	6	6	3	4	2	4	5	4	5	4	5	5		5	5	4				
NOAA ESRL O3 in-situ	2	2	2	2	3	2	5	3	1	4	4	4	4	2	5	5		5	5	4				
NOAA ESRL O3 aircraft	5	3	5	5	4	4	5	5	2	4	4	4	5	4	5	5		5	5	4				
RAOB	6	4	5	5	3	3	3	3	3	5	6	2	6	2	2	4	3	3	3	4	6	5	3	4
RBSN	6	6	5	2	2	2	2	2	2	4	6	2	6	2	2	2	3	3	4	5	6	5		
SCRIPPS C02 program	4	3	2	3	5	2	5	5	4	5	4	6	4	3	5	5		5	5	3				
SHADOZ	4	5	4	5	5	6	3	5	4	6	6	5	5	4	5	6	6	4	3	5	6	5	5	2
SKYNET	3	4	3	3	4	5	5	4	3	5	5	2	3	3	3	4	2	5	5	4		6	6	4
SMEAR	4	4	5	3	4	4	4	4	3	5	6	4	6	4	5	4	5	5	4	5				
SUOMINET	6	6	4	5	4	4	4	3	3	5	6	3	6	3	3	5	3	6	4	5	6	6	6	5
SURFRAD	3	4	5	4	5	6	5	4	3	5	6	2	6	2	5	4	3	6	5	6	5	4		
TCCON	6	6	5	5	4	5	6	4	5	3	5	4	5	6	4	5		3	5	2	3	2	5	1
TOLNET	3	2	3	4	4	3	4	3	4	3	4	5	4	4	4	5	5	5	3	3	5	2	2	1
USCRN	6	6	5	6	6	6	5	6	5	6	6	2	6	6	6	4	4	6	5	6	5	5	2	5
WOUDC	5	4	5	5	5	4	4	4	4	4	5		4	3	4	6	6	4	3	4	4	4	3	

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Figure 4. Summary of assessment results for all the networks that were assessed by GAIA-CLIM as detailed in Appendix A. Note that colour assignments follow the in-line key given in Figure 3.

of the results is given in Section 4.3, although readers interested in applicability of a given network to their application area may find it more useful to consider Figure 4 and the equivalent figures to Figure 3 from the GAIA-CLIM project website.

4.2 Robustness of assessments

The main issue in the use of the assessment tools developed is related to the inevitable and irreducible level of subjectivity involved. Even though quantifiable metrics are used and backed up by guidance (Section 3.3.1.1 provides an example), interpretation shall vary from assessor to assessor. The guidance cannot envisage all use cases and there may be ambiguity as to the most appropriate categorisation because some, but not all, criteria for a range of possible scores for a sub-category may be met simultaneously. Assessor-to-assessor uncertainty has been evaluated through a redundancy exercise based on the compilation of the matrix for the same network by at least three assessors working independently for five networks: EARLINET (Pappalardo et al., 2015), GRUAN (Bodeker et al., 2016), TCCON (Toon et al., 2009), AERONET (Holben et al., 1998), and NDACC (Vigouroux et al., 2015, NDACC, 2015)

The outcome of the exercise shows a minimum uncertainty in the attribution of the maturity matrix scores among the selected compilers of ±1. In some cases, however, the uncertainty is much larger (as large as four or five in a small number of cases), and this appears to most frequently arise either from an ambiguous or inhomogeneous interpretation of the guidance or from differences in knowledge of network details and documentation. This may, in turn, point to potential for improvements in the guidance documentation and / or compilation processes in future versions. As an example, for TCCON the maturity matrices have been filled by four different assessors, each of whom are deeply familiar with the network (Figure 5, top panel). Differences are typically small, but for some categories were substantial, in one case covering the full range from 1 to 6 (concerning an interpretative issue as to whether a document indeed did (6) or did not (1) constitute a validation report). This fostered a discussion within the TCCON community, the outcome

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access
EGU

of which has been to provide a final assessment for TCCON that represents a reasonable compromise among the four compiled matrices (Figure 5, bottom panel).

Metadata	Documentation	Uncertainty characterization	Public access, feedback and update	Usage	Sustainability	Software (optional)
Standards 6, 4, 5, 4	**Formal Description of Measurement Methodology 5, 6, 6, 3**	Traceability 6, 5, 5, 5	Access 5, 5, 5, 4	Research 5, 6, 6, 5	**Siting environment 4, 2, 3, 4**	**Coding standards 1, 5, 4, 3**
Collection level 6, 4, 6, 3	**Formal Validation Report 6, 5, 4, 1**	Comparability 5, 5, 4, 4	**User feedback mechanism 4, 4, 4, 6**	Public and commercial exploitation 6, 6, 5, 5	**Scientific and expert support 5, 5, 5, 3**	**Software documentation 4, 4, 4, 2**
File level 5, 5, 5, 5	Formal Measurement Series User Guidance 6, 5, 5, 6	Uncertainty Quantification 5, 4, 5, 5	Updates to record 4, 4, 5, 5		**Programmatic support 1, 1, 2, 4**	**Portability and numerical reproducibility 6, 5, 3, 3**
		Routine Quality Management 5, 4, 5, 3	Version control 6, 5, 6, 5			Security 1, /, 2, 1
			Long term data preservation 5, 6, 6, 4			

TCCON						
Metadata	**Documentation**	**Uncertainty characterization**	**Public access, feedback and update**	**Usage**	**Sustainability**	**Software (optional)**
Standards	Formal Description of Measurement Methodology	Traceability	Access	Research	Siting environment	Coding standards
Collection level	Formal Validation Report	Comparability	User feedback mechanism	Public and commercial exploitation	Scientific and expert support	Software documentation
File level	Formal Measurement Series User Guidance	Uncertainty Quantification	Updates to record		Programmatic support	Portability and numerical reproducibility
		Routine Quality Management	Version control			Security
			Long term data preservation			
Legend						
1	2	3	4	5	6	Not applicable

Figure 5. Range of original assessments (top panel) and final collated assessment (lower panel) for TCCON network. In the top panel assessments that differ by more than one are highlighted in bold. The four assessors are always given in the same order.

4.3 Assessment results synthesis

We now go through each of the 5 major assessment strands requested for all networks to ascertain any common strands or findings that may point to more systemic issues across many networks. This may in turn point to potential for remedial actions that can be undertaken by PIs and / or funders.

4.3.1 Metadata

Table 2 reports the frequency of occurrence for the "Metadata" category of the MSMM. Relevant international standards for metadata are assessed as having been adopted by most of the networks we have considered. Classification of file level metadata also appears to be robust throughout most of the networks and includes for most of them complete location, file level and measurement specific metadata. Conversely, collection level metadata for the majority of networks can still be improved. Such collection level metadata serves to increase discoverability and usability of whole series and would constitute relatively little work for networks to address.

Metadata sub-category	1	2	3	4	5	6
Standards	0	1	5	10	16	13
Collection level	0	2	14	10	3	16
File level	0	3	1	4	37	

Table 2. Frequency of occurrence of the maturity matrix scores for the three sub-categories (Standards, Collection level, File level) of the main category "Metadata". Note that file level does not use the score category 6 and so is blacked out.

4.3.2 Documentation

Table 3 reports the frequency of occurrence of scores for the "Documentation" category of the MSMM. A high level of maturity is assessed for the provision of a formal description of measurement methodology. Most networks provide journal papers on measurement systems with updates published in a timely fashion. Formal validation reports are available via published reports or journal papers on product validation or on inter-comparison to other instruments for most networks leading to a prevalence of high scores in this category. For the provision of formal guidance to perform measurements, documentation of some form of manufacturer independent characterization and validation is provided by the majority of the networks. However, more of the networks attain baseline or comprehensive than reference scores overall in this sub-category, highlighting it as an area where improvements could be made.

Documentation sub-category	**1**	**2**	**3**	**4**	**5**	**6**
Formal description of measurement methodology	1	2	7	4	25	8
Formal validation report	1	3	4	16	13	8
Formal measurement series user guidance	0	5	6	13	11	10

Table 3. Frequency of occurrence of the maturity matrix scores for the three sub-categories (Formal description of measurement methodology, Formal validation report, Formal measurement series User Guidance) of the main category "Documentation".

4.3.3 Uncertainty

Table 4 reports the frequency of occurrence for the "Uncertainty" category of the MSMM. Routine quality monitoring is performed at a high level by most of the networks with a clear majority assessed as meeting standards expected of reference networks. Unfortunately, other aspects of the uncertainty strand show a much more mixed message highlighting limitations across many of the networks to robustly assess and quantify uncertainty in the measurements to modern metrological (measurement science) norms and expectations. Measurement traceability is assessed as constituting a reference level only for about 50% of the selected networks. Quantification of uncertainty is also of extremely mixed maturity level among the different networks, and only a few of them can be ranked with a score corresponding to the level of a reference network. Inter-comparison and cross validation to assure measurements comparability are well established mechanisms of uncertainty quantification and validation in less than a half of the reviewed networks.

Uncertainty sub-category	1	2	3	4	5	6
Traceability	1	1	7	12	22	3
Comparability	0	2	12	13	13	5
Quantification	0	5	6	13	11	10
Routine quality monitoring	0	1	1	16	16	11

Table 4. Frequency of occurrence of the maturity matrix scores for the four sub-categories (Traceability, Comparability, Quantification, and Routine quality monitoring) of the main category "Uncertainty".

4.3.4 Public access, feedback and update

Table 5 reports the frequency of occurrence for the "Public access, feedback, and update" category. In general, it was not always easy to find detailed information about data usage which may lead to some heterogeneity in assessed scores not truly reflective of the underlying network maturity. Access to networks' public databases is high and, as such most of the networks are assessed as being at a reference level. Updates to data records are mature for most of the networks along with long-term data preservation aspects. Conversely, systematic collection of user feedback is based on a robust mechanism only for a few networks and most of them are at or below a baseline level. Control of data version and preservation of the different versions also varies hugely across the networks, with most of them assessed as at a baseline level. User feedback and preservation maturity could be increased by many networks at little to no cost and would represent adoption of best practices.

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

Public access, feedback and update sub-category	1	2	3	4	5	6
Public access / archive	3	0	0	13	9	20
User feedback mechanism	0	10	5	15	6	7
Updates to record	3	1	1	9	12	18
Version control	1	8	9	8	6	7
Long-term data preservation	0	4	5	4	25	4

Table 5: Frequency of occurrence of the maturity matrix scores for the five sub-categories (Public Access/Archive, User Feedback Mechanism, Updates to record, version control, and Long-term data preservation) of the main category "Public access, feedback, and update".

4.3.5 Sustainability

Table 6 reports the frequency of occurrence for the "Sustainability" category of the MSMM. Most of the assessed networks provide high maturity scores across the board for this category. For most of the networks, long-term ownership and rights to the site are guaranteed and the site is representative. Most of the networks offer a robust scientific support framework provided by at least two experts, which includes active instrumentation research and development being undertaken. A programmatic funding support to the network activities is ensured and not dependent upon a single investigator or funding line, with only a few networks with expectation of follow-on funding (only in one case project pending). This refers to the network wide assessment of the MSMM. Consequently, networks with heterogeneous funding structures, may have single sites which still face sustainability issues.

Sustainability sub-category	1	2	3	4	5	6
Siting environment	0	0	4	6	18	16
Scientific / expert support	0	1	5	7	34	[redacted]
Programmatic support	0	1	3	12	15	13

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access

Table 6. Frequency of occurrence of the maturity matrix scores for the three sub-categories (Siting environment, Scientific/expert support, Programmatic support) of the main category "Sustainability". Note that category six is not used in Scientific/expert support.

4.3.6 Other categories not formally considered

Two main categories were deigned optional following the first few attempts at compilation: Software and Usage. Most of the maturity matrix compilers reported being either unsure of the definitions in these two categories or not able to provide the requested information. In particular, the Software category was not always able to represent the range of practices within the networks. However, it is worth noting that the Usage category has revealed that, for most of the networks, societal and economic benefits and influence on decision-makers (including policy) of the provided data is still limited. The GAIA-CLIM activities, if successful, will increase usage for the specific case of satellite characterization.

4.4 Final assessed network status and visualisation of capabilities

Based upon primarily "Uncertainty" category scores, but also taking into account scores in the four remaining main strands considered in Section 4.3, we have categorised the networks we classified as falling into Reference, Baseline and Comprehensive categories for the purposes of downstream use within GAIA-CLIM. We shall concentrate in future work upon those classified as Reference, augmented by those classified as Baseline but with Uncertainty scores close to Reference and for which work under the work package considering metrological traceability of measurements produces new reference quality data streams.

The resulting network classifications can be used to map and visualise geographical measurement capabilities by ECV, vertical domain and measurement system maturity. As an example, in Figure 6, water vapour networks classified as Comprehensive, Baseline and Reference according to the MSMM for the category "Uncertainty" are compared, and the "global" picture of all the networks measuring water vapour is also reported. In this realisation, networks measuring the vertical profile, the full column content or at the surface have not been differentiated from one another. This figure highlights that most of the networks are classified as "Baseline" in their capability to report the measurement uncertainty. Most of

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access

the water vapour measurements are collected in the Northern hemisphere and there is a clear lack of reference measurements in the Southern hemisphere. Figure 7, that reports the same comparison for the "Documentation" category of the MSMM, provides results consistent with Figure 6. Similar maps are to be made available for all primary assessment strands and GAIA-CLIM ECVs, and an interactive mapping and visualisation tool is in the advanced stages of preparation.

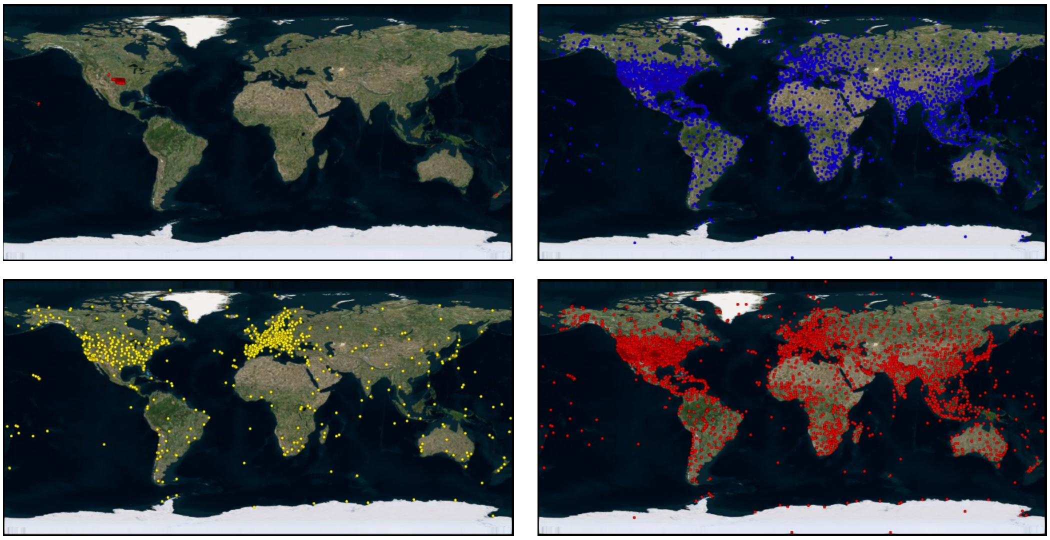

Figure 6: Classification based upon primarily "Uncertainty" MSMM category scores of the existing networks at the global scale providing in-situ water vapour measurement; upper left panel shows water vapour networks classified as "Comprehensive"; upper right panel, the networks classified as "Baseline"; lower left panel, the networks classified as "Reference"; lower right panel finally shows all the networks measuring water vapour at once. In this realisation, networks measuring the vertical profile, the full column content or at the surface have not been differentiated from one another. This figure highlights that most of the networks are classified as "Baseline" in their capability to report the measurement uncertainty and that most of the water vapour measurements are collected in the Northern hemisphere.

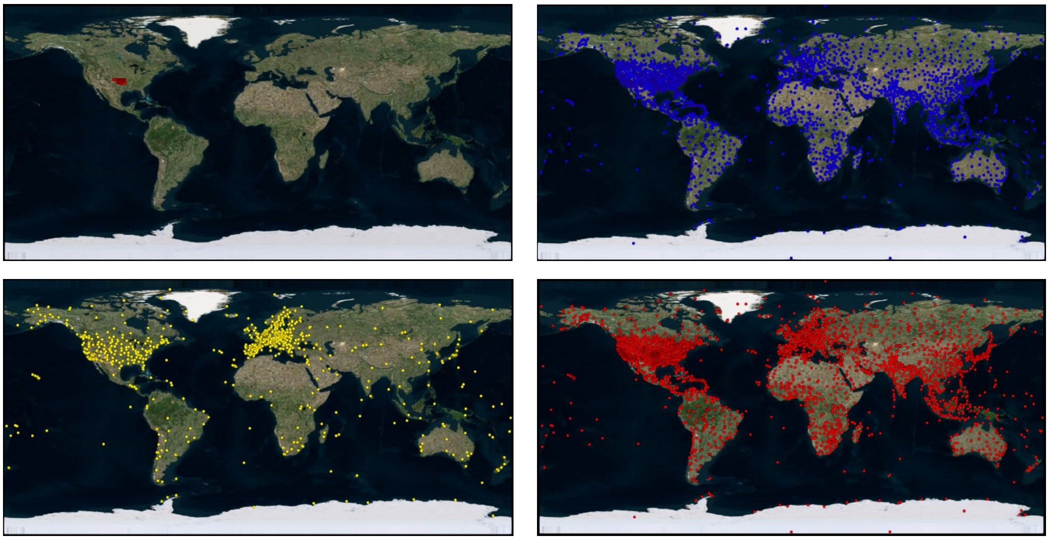

Figure 7: same as Fig. 6 but for the category "Documentation" of the MSMM

5. Discussion

5.1 Critical reflection on the exercise outcomes and limitations

Similar to the CORE-CLIMAX experience with CDR producers, there was originally a degree of user scepticism around the potential value of the assessment activity. Several networks remarked upon completion that the exercise had been useful and had led to discussions around potential innovations or improvements which could yield increased assessment scores in any future assessment but, more importantly, increase accessibility, usability and robustness of their measurement systems. This benefit was felt most strongly for those networks whose central mission is to provide the highest possible quality measurements such as USCRN (Diamond et al., 2013, Leeper et al., 2015), TCCON, NDACC and EARLINET. For example, the results of the NDACC assessment were discussed in depth at their most recent annual meeting and this led to several suggestions for improvements.

In performing an assessment at the level of the network there were recognised limitations. In the results presented in Section 4, each maturity matrix refers to the “lowest common denominator” of the performance of networks' core stations. This implies that the network assessment might not be representative of the status of the measurements performed in all the stations of the network. This implies that for networks which exhibit a substantive degree of heterogeneity in aspects of their measurement systems a further sub-selection assessment will be required by users prior to any data analysis.

Furthermore, as detailed in Section 4.2 there exists an irreducible degree of ambiguity in the performed assessments arising from assessor-to-assessor ambiguity in interpretation of guidance and / or knowledge of particular facets of the network operation. To this end, great care has been adopted by all the maturity matrix compilers to provide information as reliably as possible and in a few cases, a plenary discussion with all the network representatives has been carried out and was felt to be useful. Two categories of the MSMM, "Software" and "Usage", have been not considered robust enough at the current stage and were excluded from this assessment. They should not be adopted in future applications of the MSMM without further discussing and improving their usefulness and assessability.

After some debate amongst compilers it was agreed that for each specific assessment, the scores related to each of the sub-categories must be retained and made available. The score from the main categories only is not felt to be appropriate and does not show the real value of MSMM approach. To provide a value representative for each main category of the maturity matrix, if absolutely required, it was proposed that users must consider the minimum value from the related sub-categories to avoid providing undue confidence and also to encourage consideration of the assessment at the sub-category level.

Finally, on a practical note, people filling in a maturity matrix have provided their scores in several different ways owing to the lack of a common collection template. If the MSMM will be adopted as a tool for the self- or external assessment of a network, a new template that is able to meet the expectation of most of the compilers should be provided. An interactive online maturity matrix collection tool showing each category by clicking on the scores was considered, but its implementation was beyond the scope and the resources of GAIA-CLIM. It would, however, be a helpful development in future if broader adoption were foreseen.

5.2 Challenges to broader adoption

The MSMM approach has been used in the first instance solely for the internal purposes of GAIA-CLIM. However, there is also a broader need to articulate and adopt a system-of-systems approach, which this documentation may help to nurture (GCOS, 2014, 2015, 2016). There are significant challenges to its likely broad adoption which were recently highlighted by GCOS (GCOS, 2014), and which are expanded upon here.

5.2.1 Naming nomenclature for existing networks across and within domains

Perhaps the largest challenge is that currently a broad range of non-satellite measurement networks have been called 'Reference', 'Baseline' or 'Comprehensive' that, when assessed against the criteria detailed in Section 3, would instead fall within a different category. The lack of clarity historically regarding a system-of-systems architecture, taken together with fractured observational governance and support structures, has led to a varied use and adoption of network nomenclatures and practices both across, and within, Earth Observation science disciplines. This means that what different sub-communities concerned with environmental measurements refer to as 'Reference', 'Baseline' or indeed 'Comprehensive' network measurements is not always the same. Often it is not even remotely similar.

If a system-of-systems approach is to be broadly adopted, significant further work is required to reconcile the disparate approaches to network designations, and to manage the transition to a more trans-disciplinary approach to network assignations. There are several risks and/or challenges in any such transition:

1. National or international funding support for a measurement program may be tied to its present designation. There is a risk in enforcing any change that the funding support for the program is endangered. An example is the Ocean reference network which is not a reference network in the sense advocated here, but rather closer to Baseline capability. Nevertheless, this is still the best set of ocean observations available, and risking its loss would be a significant mistake.
2. Users may use a measurement program because of its current designation, and may get confused if measurement programs are reassigned or renamed without adequate consultation or justification.
3. The observers undertaking the measurement program may not fully understand the implications if updates to protocols and/or practices are required.

On the other side to these concerns is that allowing the status quo to continue means that users referring to e.g., a ‘Reference’ network in the marine, atmospheric and composition communities (just as by way of an example) may be comparing measurement programs that are widely differing concerning their fundamental measurement characteristics and qualities and, therefore, suitability for a given application. The status quo places the responsibility of understanding the measurement systems and networks, on a system-by-system and even ECV-by-ECV basis, firmly on the end-user. Experience shows that end-users are, understandably, unlikely to have either the time or the necessary in-depth knowledge and/or expertise to fully understand the distinctions that may exist between similarly named programs and assume, often incorrectly, that they are equivalent. This is a barrier to the effective usage of existing EO capabilities by scientists, policy makers and other end users, and will continue to be so unless and until a more holistic approach, such as suggested herein, is adopted.

5.2.2 End-User Adoption

It is clear that alongside adoption and designation of a system-of-systems framework, it is necessary to provide material to aid users to understand what the layers mean, and to show real case examples of how they can be used. GAIA-CLIM will, through its work packages, provide case study examples in the domain area of satellite measurement characterisation. However, further examples in other domain areas

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

and application areas are necessary, that will be beyond the remit of GAIA-CLIM. The MSMM shall be repeated in the new Horizon 2020 INTAROS project in 2018-19 where its use shall expand to other domains and with an Arctic region focus. The new 2016 GCOS Implementation Plan (GCOS, 2016) has an action that alludes to application of this or a modified version hereof to multiple domains (Figure 8).

Action G13:	**Review of ECV observation networks**
Action	For all ECV products not covered by a review following actions G11 and G12: develop and implement a process to regularly review ECV observation networks, comparing their products with the ECV product requirements; identify gaps between the observations and the requirements; identify any deficiencies and develop remediation plans with relevant organizations; and ensure the data is discoverable and accessible. This action may also contribute to the definition of reference grade observing network and standards The GCOS science panels should identify stakeholders who will perform this review and regularly check all ECV products are being reviewed.
Benefit	Increase quality and availability of climate observations.
Who	Organizations listed in Annex A.GCOS Panels to maintain oversight.
Time-frame	Develop and demonstrate review process in 2017. Review each ECV's observing systems at least every 4 years.
Performance Indicator	Reports of results of ECV reviews produced by panels each year.
Annual Cost	100k-1M US$, Also part of work of panels

Figure 8. Action G13 of the GCOS Implementation Plan (GCOS, 2016) as adopted by UNFCCC, which alludes to a capabilities based assessment of measurement assets which the present work may contribute to.

5.2.3 Realising technological and scientific benefits of a system-of-systems approach

Even if the layer designations and criteria documented herein were adopted, there would remain the challenge of ensuring linkages between the different components of the global observing system to realise benefits. This includes aspects such as infrastructure co-location, intercomparison campaigns, information sharing, training and development, etc.. Such inter-linkages will become both more obvious and more realisable if a system-of-systems architecture approach and assessment is adopted. Some subset of these aspects that touch upon satellite calibration/validation are covered within the regularly updated Gap Assessments and Impacts Document of GAIA-CLIM (www.gaia-clim.eu/page/gap-reference-list).

6. Summary

We have provided a proposed definition of observing system layers in a system-of-systems context and a means by which to assess in a quantifiable and objective manner demonstrable aspects of a given measurement series that help to place it into the appropriate layer. The assessment closely mirrors, but is distinct from, existing efforts to assess maturity of CDRs. In practice, the application to atmospheric ECVs will inform work within GAIA-CLIM in the creation of tools and products to be served via a Virtual Observatory facility of co-location match-ups between satellite measurements and selected non-satellite series that were assessed herein as sufficiently mature. The approach developed should be more broadly applicable to other domains and problems and if broadly adopted may have tangible benefits for data users and data providers alike. However, as this was a first attempt at such an exercise, there are undoubtedly potential improvements that could be made were it to be taken forwards. We hope that this paper provides a basis for further discussions and refinements.

Data availability

The assessment results are made available at http://www.gaia-clim.eu/page/maturity-matrix-assessment

Author contributions

PWT led the production of the assessment guidance and drafting of this paper. FM led the assessment process. FM, JS, TO, BI, ACM and GP contributed to the drafting of the guidance for assessment. KK and MdM provided a review of this guidance. MR, ET, AA, MB, MdM all contributed to the assessment of the guidance. PWT, RD and CV provided reviews and suggestions on the assessment. All authors contributed to the drafting of the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Horizon 2020 grant GAIA-CLIM, Grant no. 640276. This work benefitted from the input of D. Tan (formerly of ECMWF, currently unaffiliated) during early drafting stages. This work has been based upon the substantial work undertaken by CORE-CLIMAX and a number of precursor studies assessing dataset maturity. Without these preceding efforts this work would not have been possible. A. Apituley, G. Bodeker, B. Goodison, M. Bourassa, and Ge Peng provided feedback based upon early drafts of GAIA-CLIM (2015) that served to improve the Guidance. We gratefully acknowledge all the PIs and data managers of the measurements networks who facilitated our work to deal with the maturity matrix assessment of each of these components of the global observing system. In non-alphabetical order: A. Shimizu (NIES), M. Fujiwara (JMA), M. Palecki (NOAA), S. Lolli (NASA-JCET), T. Leblanc (NASA-JPL), W. Steinbrecht (DWD), J. Hannigan (NCAR), T. Blumenstock (KIT), N. Kaempfer (UBern), M. Campanelli (ISAC-CNR), A. Thompson (NOAA), V. Estelles (CSIC), R. Dirksen (DWD), A. Comeron (Univ. Politecnica de Catalunya), M. Fujiwara (Hokkaido University), D. Sisterson (DOE-ARM, US), T. Eck (NASA-GSFC). We also acknowledge the requirements of the BING maps add-in, its use is in agreement with the terms and conditions and privacy statement.

References

Bates, J. J. and J. L. Privette, (2012), A maturity model for assessing the completeness of climate data records, Eos Trans. AGU, 93(44), 441.

Boumans, M., (2015), Science Outside the Laboratory: Measurement in Field Science and Economics, Oxford Un. Press, ISBN 978-0-19-938828-8.

Bodeker, G., S. Bojinski, D. Cimini, R. Dirksen, M. Haeffelin, J. Hannigan, D. Hurst, T. Leblanc, F. Madonna, M. Maturilli, A. Mikalsen, R. Philipona, T. Reale, D. Seidel, D. Tan, P. Thorne, H. Vömel, and J. Wang, 2016, Reference Upper-Air Observations for Climate: From Concept to Reality. *Bull. Amer. Meteor. Soc.,* **97**, 123–135, doi: 10.1175/BAMS-D-14-00072.1.

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014, The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. Bull. Amer. Meteor. Soc., 95, 1431–1443.

Diamond, H.J., T.R. Karl, M.A. Palecki, C. Baker, J.E. Bell, R.D. Leeper, D.R. Easterling, J.H. Lawrimore, T.P. Meyers, M.R. Helfert, G. Goodge, and P.W. Thorne, 2013: U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. Bull. Amer. Meteor. Soc., 94, 485–498, doi: 10.1175/BAMS-D-12-00170.1.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H., 2014, Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4463-4490, doi:10.5194/amt-7-4463-2014

GAIA-CLIM, 2015, Report on system of systems approach adopted and rationale, http://www.gaia-clim.eu/system/files/workpkg_files/640276_Report%20on%20system%20of%20systems%20approach%20adopted%20and%20rationale.pdf

GAIA-CLIM, 2016, Report on data capabilities by ECV and by system of systems layer for ECVs measurable from space, http://www.gaia-clim.eu/system/files/workpkg_files/D1.6%20Report%20on%20data%20capabilities%20by%20ECV%20and%20by%20systems%20of%20systems%20layer.pdf

GCOS, 2014, Workshop on the review of the GCOS Surface Network (GSN), GCOS Upper-Air Network (GUAN), and related atmospheric networks, *Ispra, Italy, April 2014*, GCOS-182 [http://www.wmo.int/pages/prog/gcos/Publications/gcos-182.pdf]

GCOS, 2015, Status of the Global Observing System for Climate, GCOS-195 [http://www.wmo.int/pages/prog/gcos/Publications/GCOS-195_en.pdf]

GCOS, 2016, The Global Observing System for Climate: Implementation Needs., GCOS-200 [http://unfccc.int/files/science/workstreams/systematic_observation/application/pdf/gcos_ip_10oct2016.pdf]

GEO, 2016, GEO Strategic Plan 2016-2025 Implementing GEOSS, http://www.earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf (accessed 7/4/17)

Holben B.N., T.F. Eck, I. Slutsker, D. Tanré, J.P. Buis, A. Setzer, E. Vermote, J.A. Reagan, Y. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, 1998: AERONET - A federated instrument network and data archive for aerosol characterization, Rem. Sens. Environ., 66, 1-16.

Immler, F. J., J. Dykema, Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H. , 2010, Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Measur. Tech. **3**(5): 1217-1231

JGCM, 2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement [www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf].

Leeper, R. D., Rennie, J., and Palecki, M. A.: Observational Perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison, J. Atmos. Ocean. Tech., 32, 703– 721, 2015.

National Academy of Sciences (NAS), 2009, 'Observing weather and climate from the ground up: A nationwide network of networks.' http://dels.edu/Report/Observing-Weather-Climate-from/12540

NDACC, 2015, NDACC newsletter, http://www.ndsc.ncep.noaa.gov/news/archives/nl2015-8.pdf (accessed 7/4/17)

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmos. Meas. Tech., 7, 2389-2409, doi:10.5194/amt-7-2389-2014, 2014.

Schulz, J., V. John, A. Kaiser-Weiss, C. Merchant, D.Tan, E. Swinnen and R. Roebeling, 2017, European Climate Data Record Capacity Assessment, in prep for Geoscience Data Journal

Seidel, D. J., et al., 2009, "REFERENCE UPPER-AIR OBSERVATIONS FOR CLIMATE Rationale, Progress, and Plans." Bulletin of the American Meteorological Society **90**(3): 361-369.

Toon, G. C., J.-F. L. Blavier, R. A. Washenfelder, D. Wunch, G. Keppel-Aleks, P. O. Wennberg, B. J. Connor, V. Sherlock, D. W. T. Griffith, N. M. Deutscher, and J. Notholt (2009), Total Column Carbon Observing Network (TCCON), in Fourier Transform Spectroscopy, pp. 4-7, Optical Society of America. Available from: http://www.opticsinfobase.org/abstra...=FTS-2009-JMA3

Vigouroux, C., Blumenstock, T., Coffey, M., Errera, Q., García, O., Jones, N. B., Hannigan, J. W., Hase, F., Liley, B., Mahieu, E., Mellqvist, J., Notholt, J., Palm, M., Persson, G., Schneider, M., Servais, C., Smale, D., Thölix, L., and De Mazière, M.: Trends of ozone total columns and vertical distribution from FTIR observations at eight NDACC stations around the globe, Atmos. Chem. Phys., 15, 2915-2933, doi:10.5194/acp-15-2915-2015, 2015.

WIGOS, 2015a, WMO technical regulations (WMO-No. 49) - manual on wigos

WIGOS, 2015b, WMO technical regulations (WMO-NO. 49) - manual on wigos attachment wigos metadata standard

WMO / BIPM, 2010, Measurement Challenges for Global Observation Systems for Climate Change Monitoring: Traceability, Stability and Uncertainty, WMO/TD-No. 1557, Rapport BIPM-2010/08

WMO, 2015, WMO-No. 1157. Seventeenth World Meteorological. Congress. Geneva. 25 May– 12 June 2015.

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Appendix A.

Table summarising pertinent details of those networks considered in Section 4, listing those 49 networks reviewed within GAIA-CLIM Task 1.2 for which complete discovery metadata have been collected. The first column reports the measurement domain, the second the network acronym, the third, the network coverage, the fourth includes the number of measured ECVs (repeated for columns 5 to 8). Those which maturity assessments were performed for and which are discussed in Section 4.1 forwards are italicised.

Domain	**Network**	**Coverage**	**ECVs**	**Domain**	**Network**	**Coverage**	**ECVs**
Atmosphere	*ACTRIS*	*Regional*	*Aerosols, NOx, VOCs*	*Atmosphere*	*AD-Net*	*Regional*	*Aerosols*
Atmosphere	*AERONET/ PHOTONS*	*Global*	*Aerosols*	*Atmosphere*	*AGS-Net*	*Global*	*Aerosols*
Atmosphere	AGAGE	Global	Aerosols	*Atmosphere*	*AMeDAS*	*Regional*	*Temperature*
Ocean	*ARGO*	*Global*	*Temperature Salinity*	*Atmosphere*	*ARM*	*Regional*	*Numerous*
Atmosphere	*BSRN*	*Global*	*Radiation Aerosols*	Atmosphere	CAPMoN	Regional	Aerosols NOx, O_3
Atmosphere	CARSNET	Regional	Aerosols Water vapour	Atmosphere	CASTNET	Regional	Aerosols O_3
Atmosphere	CAWNET	Regional	Aerosols	Atmosphere	CREST	Regional	Aerosols
Atmosphere	EANET	Regional	Aerosols O_3	*Atmosphere*	*EARLINET*	*Regional*	*Aerosols*
Atmosphere	*EMEP*	*Regional*	*Aerosols*	*Atmosphere*	*EPA*	*Regional*	*Aerosols*
Atmosphere	*ESRL*	*Global*	*Aerosols* CO_2, *CO,* CH_4, *VOCs*	*Atmosphere*	*EUREF*	*Regional*	*Water Vapour*
Atmosphere	*EuroSkyRad*	*Regional*	*Aerosols*	Land	FLUXNET	Global	Albedo, CO_2 Water Vapour
Atmosphere	GAW GALION	Global	Aerosols	*Atmosphere*	*GAW PFR*	*Global*	*Aerosols*

Geosci. Instrum. Method. Data Syst. Discuss., doi:10.5194/gi-2017-29, 2017
Manuscript under review for journal Geosci. Instrum. Method. Data Syst.
Discussion started: 11 April 2017

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access

Atmosphere	GPS-Met	Global	Water Vapour	Atmosphere	GRUAN	Global	Temperature Water Vapour
Atmosphere	GSN	Global	Temperature Precipitation	Atmosphere	GUAN	Global	Temperature Water Vapour
Atmosphere	ICOS	Global	CO, CO_2, CH_4	Atmosphere	IDAF	Regional	Aerosols, O_3, NO_2
Atmosphere	IGS	Global	Water Vapour	Atmosphere	IMPROVE	Regional	Aerosols
Atmosphere	LALINET-ALINE	Regional	Aerosols	Atmosphere	MESONET	Regional	Temperature, Water Vapour
Atmosphere	MPLNET	Global	Aerosols	Atmosphere	MWRNET	Global	Temperature, Water Vapour
Atmosphere	NDACC	Global	Various	Atmosphere	NPS	Regional	Aerosols
Atmosphere	RAOB	Global	Temperature, Water Vapour	Atmosphere	RBSN	Global	Temperature, precipitation, pressure
Atmosphere	Scripps CO_2	Regional	CO_2	Atmosphere	SHADOZ	Regional	O_3
Atmosphere	SKYNET	Regional	Aerosol	Atmosphere	SMEAR	Regional	Temperature, Water Vapour
Atmosphere	SUOMINET	Global	Water Vapour	Atmosphere	SURFRAD	Regional	Radiation, aerosols
Atmosphere	TCCON	Global	CO_2, CH_4, CO	Atmosphere	TOLNET	Regional	O_3
Atmosphere	USCRN	Regional	Temperature, precipitation, Water Vapour	Atmosphere	WOUDC	Global	O_3