Proposal for formalisation and standardisation of tiered network approach across domains and observing system programs

Purpose

To propose an approach for the tiering of observing networks, in order to specify the purpose and classify the quality and characteristics of observing networks within and across domains, and to provide a hierarchy of data usage in support of user application areas.

A tiered network approach

Not all observations are equal and nor do they need to be. Observation programmes are run by a diverse range of stakeholders, within a very distinct set of research and/or operational guidelines, to meet specific customers' needs. Data quality needs are inherently linked to intended applications. Some applications may require very few observations (even only one) of the highest quality, whereas other applications may require a large number of observations but be able to accept lower observation quality. Every possible combination exists in-between these extremes. To make sense of this mosaic of available observations, what is critical is that users, who cannot reasonably be expected to have in-depth knowledge of every single observation programme that exists, are guided to use the most appropriate set of observations to meet their specific needs.

In the absence to date of an agreed global governance to address WIGOS Observing Network Design Principle No. 7 (Designing through a tiered approach[1]) a range of naming conventions and assessment criteria have arisen within different domains to try to make sense of heterogeneous observing capabilities.

To date, this has had varying degrees of success and heterogeneity in application has led to issues. For example, the same terms in different contexts are used to denote non-equivalent capabilities which can be highly misleading to users. This application of heterogeneous definitions and processes for assessments, including the granularity of application, severely inhibits data interoperability and optimal use of observations both within and across Earth system domains.

However, when considered objectively, there are fundamental characteristics of the observation process that irrespective of the domain or observing technique all contribute to data quality and useability. These characteristics include, but are not limited to:

- Commitment to and expectation for long-term continuous data sets
- Commitment to maintained quality
- Documentation practices
- Data communication, handling and storage
- Data exchange and data policy
- Quality of instrumentation and methods of observation
- Degree of traceability to SI or accepted community standards
- Data uncertainty characterisation
- Siting suitability and stability
- Metadata provision and retention

[1] See WMO-No. 1160m

Given that observations and the resulting data are and will continue to be heterogeneous in regard to the above characteristics, the use of tiered network designations using, but not limited to, these characteristics to determine tier membership is desirable from management, suppliers, operations, and user perspectives. It will allow the composite observing systems to evolve in a cost-effective manner based on the best mix of observing stations distributed over the tiers, recognising and realising synergies between tiers. Assignment to tiers would also highlight critical gaps in capability which stakeholders may then seek to remedy.

The tiering concept can, in theory, be applied at various levels of granularity from network-wide designations down to a consideration of individual observation series at individual stations. Increased granularity of assessment implies increased efforts required to undertake and then maintain assessment and the relative costs and benefits of different granularity of assessment must be carefully considered. As a first step, it is proposed that the initial implementation of observing system tiering be performed for networks, consistent with the articulated guidance in existing WIGOS regulatory materials that speaks of network designations[2]. Subsequent applications at finer granularity (stations or individual observations) should be considered in future.

Therefore, the proposal to INFCOM is for WMO to undertake a programme to create a set of unified tiered network definitions applicable within and across domains, followed by designation and governance, to enable greater data interoperability and improved exploitation of both existing and new observation programmes.

If this concept is accepted, then the next steps would be to define and adopt:

1) A set of tiers with definitions that are strongly linked to existing mechanisms within WMO, for example OSCAR, and taking into consideration user community use cases.

2) A set of tier assessment criteria that are strongly linked to existing mechanisms within WMO, for example the WMO Siting Classification Scheme, and the WMO Measurement Quality Classification Scheme (MQCS).

Such an assessment would use objective criteria against assessable aspects of network measurement programmes such as those listed previously to decide the level of maturity of the network observational programme. A final assessment may be made based on collective performance across a range of categories.

Consideration would also be required as to how to make network designations visible to users (e.g., via OSCAR/Surface) to enable systematic exploitation by users.

Although initially scoped for in-situ and ground-based remote sensing (sub-orbital) capabilities, the tiered networks concept is equally applicable to remote sensing and satellite observations, particularly as the satellite observing system becomes increasingly heterogeneous[3].

Benefits

Specific benefits have been categorized in three main classes: benefits for users, benefits for verification and validation, and benefits for providers.

In terms of benefits for **users**, a tiered network structure enables optimal use and improved understanding of observations, by easily conveying the quality, reliability, and accessibility of a station's measurements. Tiering leads to trickle-down innovation, with higher tiers helping to make scientific sense of lower tiers (as evidenced in GRUAN). This approach will also allow the users to bridge scales of observations and potentially fill gaps. Higher quality data will improve

[2] **7. Designing through a tiered approach Observing** network design should use a tiered structure, through which information from reference observations of high quality can be transferred to other observations and used to improve their quality and utility.

[3] Traditionally the majority of space-based observations were taken either by Members via dedicated meteorological mission satellites or by space agencies flying EO missions. With new absolute calibration missions (TRUTHS, CLARREO etc.) and the advent of commercial / quasi-commercial constellations the space component is becoming increasingly heterogeneous and may similarly benefit from some form of tier designation with similar design principles and naming conventions to aid interoperability with the non-space segment.

the ability to characterize NWP and climate model bias and uncertainties. If a sufficient number of top tier stations with consistent historical records are available, that will allow calibration/validation of long-term stability reanalysis products, which is essential in discerning whether any trend or discontinuity in a reanalysis product is a genuine climatic signal or is an artefact.

The benefits of reference level for climate applications and for **verification and validation** of remaining observations, including the space segment**,** have already been recognized. Well-defined and supported reference stations/networks will naturally lead to better support for remote-sensing system validation/calibration. Analogously high quality, reference space-based observations can be used to verify and validate surface-based observations. NWP-generated "fields" (temperature, humidity, precipitation, etc) will also benefit from a quantified reference and baseline set of tiered measurements for verification/validation, leading to more accurate modelling and predictions.

Benefits for **providers** are mostly in the optimization, cost-effectiveness, and design and planning of networks. Through this approach, valuable and scientifically defensible information will be provided to network designers and operators so they may inform their decision-making within their constraints. This will promote an optimized network design and optimization of network sites. The tiered approach enables optimization of investment and provides a rational basis for funding agencies to assess trade-offs with clearly articulated defensible minimum fit for purpose investment and spending. Finally, at the station level it will give funding agencies the motivation to support stations of a higher tier; encourage station owners to improve their observational capabilities in various ways (e.g. better uncertainty assessments) in order to meet the requirements of a higher tier; and increase the visibility, accessibility, and use of stations that produce usable lower quality data but do not fully comply with WMO or other standards.

An Initial Concept for Tiering

The proposition is to have following three tiers, Reference, Baseline, Additional, supplemented by a 4th tier Ancillary for networks not belonging to the first 3 tiers (useful observations that do not fulfil the requirements of the first three tiers, i.e. poor metadata, unknown/heterogeneous ownership etc.) or which could not be assessed, called Ancillary / Unclassified.

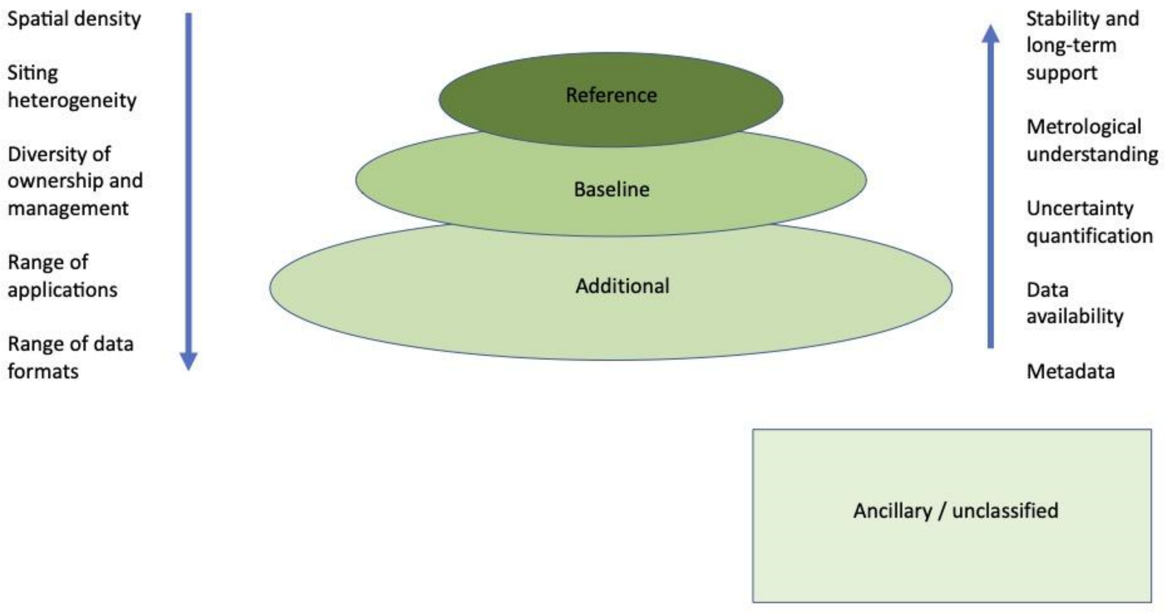

Figure 1. *Illustration of the concept of tiers and denoting a subset of key facets of the networks and how these may either increase in going from reference to lower tiers (left hand side) or vice-versa (right hand side).*

The characteristics to be met by the networks to belong to the different tiers must be clearly defined with explicit guidance. Networks should be qualified by quality and performance criteria and independently of considerations of ownership, as there are third party (non-NMHS) operated networks and observatories that are of very high quality and can objectively be assigned to the appropriate tier. The tiered approach reflects that there is, in general, a trade-off between spatial-temporal observational density and aspects of observing quality. Sites delivering observations of high-quality are expensive to maintain and often require substantive technical staffing and are therefore expected to be relatively few in numbers. Geographic completeness and representativeness in terms of sampling and resolution/coverage can be obtained with sites that are of lower quality. The system of systems that combines such components in a seamless manner would ensure optimal use of observational capabilities. Specifically, higher quality tiers can provide observational data of high quality that can be used to make sense of the remaining data in lower tiers. Conversely, observations taken at lower tiers provide the rich spatial-temporal detail necessary to understand real geophysical signals and gradients that sparse networks of high-quality observations cannot hope to capture.

The first required step is to define the characteristics of each 'tier'. The tiers are delineated by demonstrable measurement qualities such as: traceability, uncertainty quantification, metadata, comparability[4], data completeness and continuity, documentation, record longevity (where appropriate), measurement programme stability and sustainability.

The proposal would be to, at least initially, designate this on a network basis not an individual observing stream basis. Many networks are constituted by multiple instruments which may have distinct maturity. Oftentimes this may be for good reason. For example, to understand effects on the principal measurement it may be necessary to observe several additional quantities of interest. However, these may not need to be observed to the same level as the principal measurement series. It would be incredibly challenging to disentangle such cases. Furthermore, from an ease of application perspective it makes a lot of sense to, at least in the first instance, consider these designations at the network level. This does not preclude, at a later stage, undertaking such an assessment at the finer granularity of individual measurement series. Here the network is proposed to be defined as belonging to a national or international collection of measurements with a common designation and / or a common governance / aim.

The definition of the first 3 tiers from the characteristics as proposed in GAIA-CLIM, 2015 are suggested to be used as a starting point. These are as follows:

Reference observing networks: These networks provide metrologically traceable observations, with quantified uncertainty, at a limited number of locations, or for a limited number of observing platforms, for which traceability has been attained. These measurements aim to meet the 'goal requirements' in the WMO OSCAR requirement database.

- 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 Metrology Institutes), using accepted best practices at any time as documented in the accessible literature.
- 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.
- The measurements are representative of the broader surroundings and the site environment is stable (e.g. there is no ongoing or planned large-scale land-use or land-cover change of human origin for terrestrial sites).

[4] Here comparability is used in the sense that two sets of observations are able to be compared to one another. This comparison should show the way how two sets of measurements differ from one another because they measure distinct geophysical states rather than because the two systems differ in ways that impacts the measurement series in systematically distinct manners

- 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 by any user.
- The measurement and its uncertainty are verified through complementary, redundant observations of the same measurand on a sufficiently routine basis which might vary by instrument and site type. This can be achieved via a variety of means including sustained redundant observations, round-robin intercomparisons, travelling standards.
- All data and metadata are made freely available to users.
- The observations programme is actively managed and has a commitment to long-term operation, to the extent possible.
- Change management is robust including a sufficient programme of parallel and/or redundant measurements to fully understand the impact of any changes that do occur upon the measurement uncertainties. 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.

Current potential examples: GCOS Reference Upper Air Network. GAW global stations, Cryonet clusters, Ocean sites, ARGO.

Baseline observing networks: These networks provide long-term records that are of a sufficient spatial density to be capable of characterising regional, hemispheric and global-scale features. They aim to meet, as a minimum, the "threshold" level in the WMO Observing Systems Capability Analysis and Review Tool (OSCAR) requirements database.

- The baseline network is a globally, regionally and nationally representative set of observations capable of capturing, at a minimum: global, hemispheric and continental scale changes and variability. As such, a baseline network may be considered a minimum and highest priority subset of the Additional networks, which should be actively curated and retained.
- The measurements are periodically assessed, either against other instruments measuring the same geophysical parameters at the same site or, alternatively / in addition, through intercomparison campaigns held under international or national auspices. These activities provide understanding of the relative performance of different techniques in use. Ideally, such intercomparisons should include reference quality measurements / networks, to realise scientific benefits.
- Representative uncertainties, that are based upon understanding of instrument performance or peer reviewed lines of evidence, are available.
- Metadata about changes in observing practices and instrumentation are retained.
- All data and key metadata are made freely available to users.
- The observations have a long-term commitment.
- Changes to the measurement programme 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.
- The measurements aim to meet stakeholder stated requirements.

Current potential examples: GBON, RBON, GUAN, GSN, GAW contributing networks, Cryonet sites, Floating buoy program.

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

- The additional 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 some regional and national operational observing networks and 3rd party data.
- Representative uncertainties based upon e.g. instrument manufacturer specification and knowledge of operations should be provided. In their absence gross uncertainties based upon e.g. expert or operator judgement should be provided.
- Metadata should be retained.
- All data and key metadata should as much as possible be made freely available to users
- Although encouraged, long-term operation is not required.

Current potential examples: GOS, Some mesonets, VOS network, Supplemental radiosondes, commercial aircraft.

Ancillary / Unclassified observing networks: These networks provide data information for which no information is available concerning the quality of these stations/networks (where networks may be a term used very loosely e.g. collections of private citizen observations, car sensors, signal attenuation delays of mobile telemetry etc.)

- Networks with insufficient amount of information to carry out a quality assessment needed for the assignment to one of the 3 tiers. (e.g. those third Party and citizen science observations with insufficient provenance).
- Includes stations or networks that have not been assessed or cannot meet the standard of the Additional Tier.

Current potential examples are car sensors and citizen science observations.

An example of an assessment approach for assigning networks to tiers

Similar work has already been done in the past and can serve as an example. An assessment on how different networks can be assigned to different tiers was proposed in GAIA-CLIM[5]. Their work concentrated on networks measuring primarily a subset of the atmospheric Essential Climate Variables of interest to GAIA–CLIM activities. However, the proposed approach could probably be more widely applied across a broad range of application areas and modified to account for the agreed tiers herein and to be acceptable across all domains and to Members. Figure 2 provides an example using the GAIA-CLIM assessment criteria for the GCOS Reference Upper Air Network as it was configured in 2017.

[5] https://gi.copernicus.org/articles/6/453/2017/gi-6-453-2017.pdf

GRUAN						
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 reproducability
		Routine Quality Management	Version control			Security
			Long term data preservation			
Legend						
1	2	3	4	5	6	Not applicable

The maturity assessment carried out above was performed under the auspices of GAIA-CLIM, http://www.gaia-clim.eu/page/maturity-matrix-assessment, in September 2016. It assesses certain quantifiable aspects of typical measurement system maturity across the network for those ECVs and associated measurement systems that are relevant to GAIA-CLIM ,www.gaia-clim.eu.
Users should be aware that this is a first effort to systematically quantify measurement system performance. Redundant assessments suggest a minimum uncertainty arising from assessor-to-assessor variations in any category of at least 1 score. Although the assessment may be useful to certain applications; at this time and until it is more broadly tested it should not constitute a primary or unique decision-making tool.

Figure 2. Example of the maturity matrix assessment approach developed in GAIA-CLIM as applied to the GCOS Reference Upper Air Network in 2017. Higher scores denote higher quality / maturity. Some categories only extend to a score of 5.

Next steps

It is necessary first to ascertain whether the proposed overarching concept makes sense and is acceptable to a broad range of stakeholders including, but not limited to, Members.

If the concept has broad acceptance, then the task team would be reconstituted, to consult, develop and agree upon a set of criteria to be used to objectively assign candidate networks to appropriate tiers. These criteria will need to be developed in a manner that is workable across domains, WMO programmes, and ideally with broader applicability outside WMO.

A mechanism of governance for the process that represents a sustainable approach for INFCOM and Members would need to be developed along with how to integrate the approach into regulatory materials, OSCAR/Surface (and perhaps wider OSCAR) and WDQMS processes.

Once completed to the satisfaction of INFCOM the above would then need to be approved by Members at Congress prior to implementation.

____