WMO/IOC/UNEP/ICSU
GLOBAL CLIMATE OBSERVING SYSTEM (GCOS)

Doc. 2.4
(18.II.2013)

5th GRUAN Implementation-Coordination Meeting (ICM-5)
De Bilt, Netherlands
25 February – 1 March 2013

Session 2

The GRUAN Guide to Operation v1.1.0.2 (DRAFT)

(Submitted by Greg Bodeker)

Summary and Purpose of Document

The GCOS Reference Upper-Air Network (GRUAN) guide provides both mandatory operating requirements and guidelines on how to achieve the operating protocols specified in the GRUAN Manual.

THE GRUAN GUIDE TO OPERATIONS

Version 1.1.0.2

Purpose of this Guide

The GCOS Reference Upper-Air Network (GRUAN) guide provides both mandatory operating requirements and guidelines on how to achieve the operating protocols specified in the GRUAN Manual. Mandatory operating protocols are distinguished by the words 'shall' or 'must' while guidelines are distinguished by the words 'could' or 'should'. The primary goals of GRUAN are to provide vertical profiles of reference measurements suitable for reliably detecting changes in global and regional climate on decadal time scales, initially for temperature, pressure and water vapour, with the aim of expanding to other essential climate variables (ECVs), as resources permit, in liaison with other existing scientific networks. The measurements will provide a traceable reference standard for global satellite-based measurements of atmospheric ECVs, will ensure that potential gaps in satellite measurement programmes do not invalidate the long-term climate record, will be a reference standard for the measurements made within the existing GCOS Upper-Air Network (GUAN), and will provide data to fully characterize the properties of the atmospheric column. These goals have been agreed to by GCOS (Global Climate Observing System) and WMO (World Meteorological Organisation).

The Guide defines the requirements for GRUAN site operations, including requirements on measurement uncertainty and long-term stability of measurement time series. It establishes the philosophy under which GRUAN shall operate and informs current and future GRUAN sites of the expected *modus operandi* for GRUAN. This guide recognizes that GRUAN is a heterogeneous network that includes sites from both the research community and the operational meteorological community. A GRUAN site may be a scientific observing site outside of the WMO operational Global Observing System, but the long-term observational procedures are expected to follow the guidelines laid down in this Guide and in the GRUAN Manual whether it is primarily a scientific site or whether it is already part of the Global Observing System, and the WMO Manuals shall reference these practices. The mandatory practices required of GRUAN sites as detailed in this Guide and in the GRUAN Manual, reflect GRUAN's primary goal of providing reference quality observations of the atmospheric column while accommodating the diverse capabilities of sites within the network.

Relevant information from this Guide and from the GRUAN Manual is expected to be incorporated into the WMO Manual on the Global Observing System (WMO-No. 544) and the Guide on the Global Observing System (WMO-No. 488). Where possible, this Guide provides more in-depth detail on specific methodologies, and these shall be appropriate for incorporation into, or referenced by, the WMO Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8). The GRUAN Manual and Guide are supported by a series of technical documents listed on the GRUAN web site at http://www.gruan.org.

Table of Contents

1. INTRODUCTION

1.1. GRUAN heritage

The need for a reference upper-air network to better meet the needs of the international climate research community has long been recognized (e.g. Trenberth, 2003). In response to this need, the inception of the GCOS Reference Upper-Air Network (GRUAN; GCOS-112, GCOS-134) was formalized through a series of meetings held under the joint auspices of GCOS and NOAA between 2005 and 2007 when a reference upper-air network, envisaged to eventually include 30-40 sites worldwide, was first clearly articulated. In contrast to the GCOS Upper-Air Network (GUAN), which is based on weather observing stations, GRUAN is specifically designed for climate research. Therefore, rather than being a purely operational network like GUAN, it is a network that serves the international climate community through a combination of research and operational activities, giving high quality operational network observations and elements of research and development for the future. GRUAN provides reference observations of upper-air essential climate variables (ECVs; GCOS-138) through a combination of *in situ* measurements made from balloon-borne instruments and from ground-based remote sensors. Furthermore, management decisions in GRUAN are driven by a variety of requirements for long-term measurements of assured measurement stability, including the need for good operational practices to ensure this stability. On one hand GRUAN is partly a research network constantly striving to improve measurement techniques, and quantify and reduce measurement uncertainties by improving precision and accuracy, but on the other hand the network measurements need to be made in a stable way over multi-decadal time scales to achieve data homogeneity in time and spatially between measurement sites. These two aspects of GRUAN operations are not mutually exclusive, but do need to be carefully balanced. The dual-purpose nature of GRUAN has been accommodated in this Guide.

1.2. The purpose of GRUAN

As detailed in GCOS-112, the purpose of GRUAN is to:

i) Provide long-term high quality climate records;

ii) Constrain and calibrate data from more spatially-comprehensive global observing systems (including satellites and current radiosonde networks); and

iii) Fully characterize the properties of the atmospheric column.

To achieve these goals, sites within the network will provide vertical profiles of reference measurements of temperature, pressure and water vapour (and additional essential climate variables) suitable for reliably detecting changes in global and regional climate, on multi-decadal time scales, for major climatically distinct regions of the globe. The uniformity and coherence of standard operating procedures at GRUAN sites and the resultant homogeneity of GRUAN climate data records not only provides a global reference standard for operational upper-air network sites, but improves the detection of changes in the climate of the troposphere and stratosphere. Measurements at GRUAN sites will also provide a reference for global satellite-based measurements of atmospheric ECVs. This facilitates the creation of seamless, stable, and long-term databases of satellite-based measurements suitable for detecting trends and variability in climate in the troposphere and stratosphere. Given the importance of the satellite community as a user of GRUAN data, Section 1.6 is dedicated to discussing how GRUAN serves that community. In achieving these goals, GRUAN will ensure that any interruptions in satellite-based measurement programmes do not invalidate the long-term climate data record. GRUAN shall, where practicable (see Section 8), provide observations in near real-time (i.e. within 2 hours of the measurement) for incorporation in meteorological analysis to fulfil the requirement of providing a reference to the operational observations.

In achieving its goals, GRUAN will address some of the current deficiencies of the GUAN network. The reliable detection of the vertical structure of changes in climate variables in the atmosphere requires high quality observations of atmospheric profiles with well characterised measurement uncertainties. GUAN provides upper-air measurements over large regions of the globe using radiosondes that in many cases are similar to those used in GRUAN. However GUAN sites seldom include additional systems to validate data stability, and rely on the assumption of stability in the radiosonde quality with time. GUAN sites are also more likely to use instrument manufacturer proprietary software which does not permit a robust traceability of all sources of measurement uncertainty. If GRUAN can identify the changes that occur in production consumables, this will benefit those using GUAN measurements and all users of WIGOS (WMO Integrated Global Observing System) and GAW (Global Atmospheric Watch) upper-air measurements.

Four key user groups of GRUAN data products are:

i) *The climate detection and attribution community:* the long-term stability and homogeneity of GRUAN data provide time series needed to robustly detect and attribute changes in the climate of the free atmosphere. GRUAN data will also be used to constrain and validate data from more spatially comprehensive global networks for improved climate detection and attribution.

ii) *The satellite community:* GRUAN data products are used to validate satellite-based measurements and to provide the input needed for radiative transfer calculations required to improve and evaluate retrieval algorithms.

iii) *The atmospheric process studies community:* by providing high precision and high vertical resolution measurements with defined uncertainties of a range of upper-air climate variables, GRUAN data products will aid in developing a deeper understanding of the processes affecting the atmospheric column. Because GRUAN will make profile measurements at vertical resolutions much higher than can be retrieved from satellites, it will provide valuable insights into the potential limitations of satellite-based measurements for the analyses of specific atmospheric phenomena.

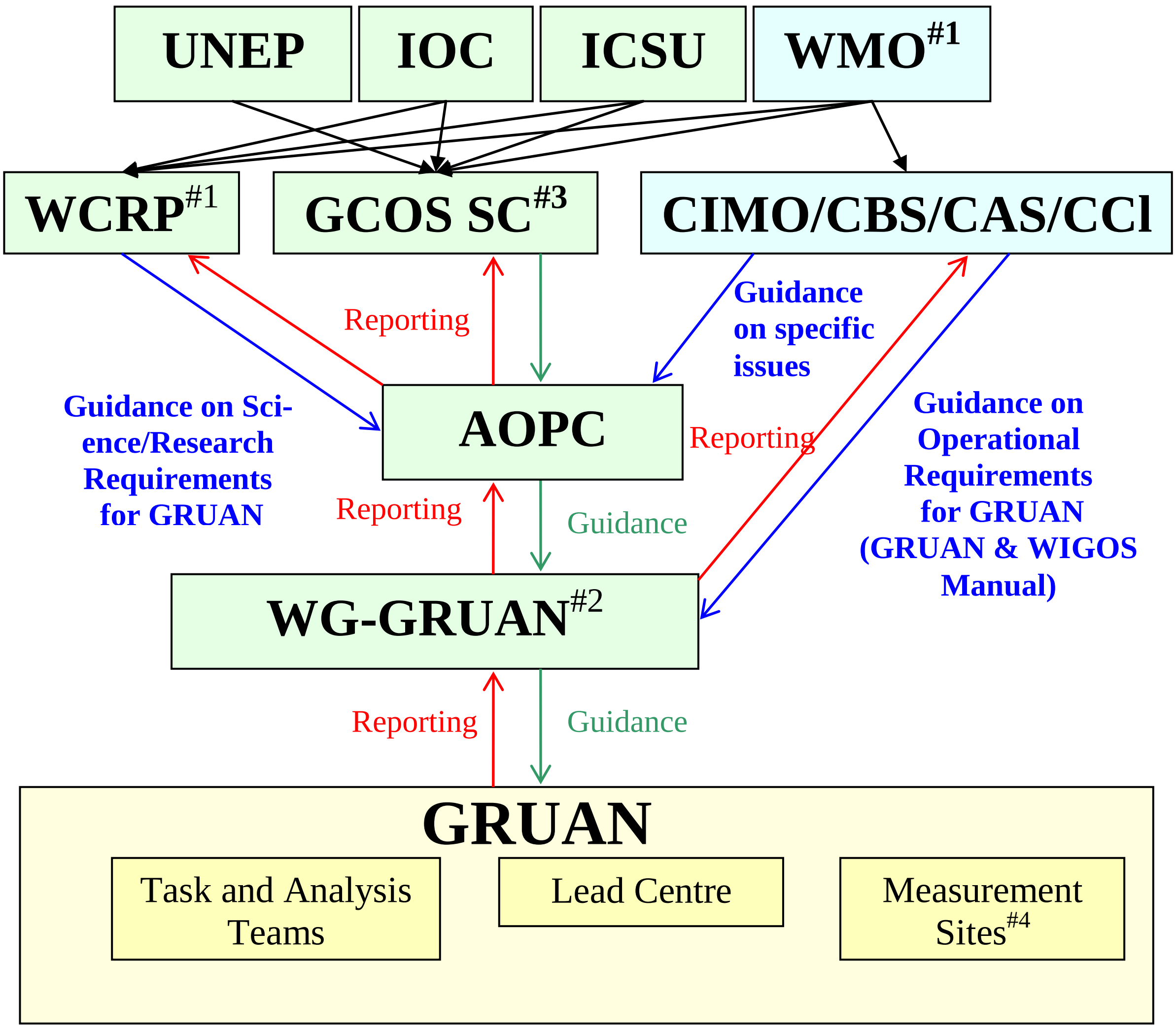
iv) *The numerical weather prediction (NWP) community:* The reference quality of GRUAN data makes them useful for verifying NWP model outputs, and for validating and correcting other data being assimilated into NWP models. Measurements made at GRUAN sites can also be directly assimilated in real-time, or near real-time, into NWP models, provided this is not detrimental to achieving the primary purposes of the network, as defined above. GRUAN reference measurements can also be assimilated into meteorological reanalyses.

In the context of the other WMO observing systems, GRUAN will need to be the climate reference backbone of the existing global operational upper-air network. As noted in GCOS-112, GRUAN sites need not necessarily be current GUAN sites. Because GUAN sites often operate with different equipment, sensors, and operating protocols, the different requirements of GRUAN and GUAN operations may require careful management. The envisaged capabilities of a fully-implemented GRUAN are detailed in GCOS-112. The scientific justification and requirements for GRUAN are summarized in Section 3 of GCOS-112 and in Seidel et al. (2009) and are not repeated here. Continued implementation of GRUAN is specifically called for under Action A16 of the 2010 update to the implementation plan for GCOS (GCOS-138). The connection of GRUAN to other global climate observation networks is detailed further below.

1.3. Organisation and design of GRUAN

GRUAN will operate under the joint governance of GCOS and WMO as a WIGOS Implementation Project. A schematic outline of the GRUAN governance structure is given in Figure 1. The GCOS Steering Committee guides the GCOS/AOPC (Atmospheric Observations Panel for Climate). The AOPC in turn guides the WG-GRUAN (Working Group on GRUAN) which provides working oversight of GRUAN and includes representatives at the working level from WMO. GCOS and WMO will select those groups (e.g. AOPC, or WMO Technical Commission working groups/experts) through which WG-GRUAN will report. The WG-GRUAN is responsible for GRUAN site selection and certification (Section 5) and is responsible for developing and approving guidelines for observations and data.

A GRUAN Lead Centre, agreed to by GCOS and WMO, will be responsible for integrating best practices into GRUAN operations, managing the network systems, and data management. This Lead Centre is currently operated by the German Meteorological Service (DWD) at the Lindenberg Meteorological Observatory in Germany. The GRUAN Lead Centre acts as the interface be-



Notes

1. WCRP identifies scientific and research requirements for GRUAN, while WMO identifies operational requirements.
2. Composition of WG-GRUAN to be determined by the AOPC in consultation with WMO and should include:
   - one representative from each of CIMO, CBS, CAS and CCl; these representatives will be responsible for reporting back to their respective Technical Commission;
   - others (according to its Terms of Reference)
3. Global Climate Observing System Steering Committee.
4. GRUAN Measurement Sites are contributed by Members of WMO.

Figure 1: Schematic outline of the structure of GRUAN.

tween GRUAN and the community of users of GRUAN products. For example, data transfer to end-users is not made from GRUAN sites but is first shared within the GRUAN community, subjected to the QA/QC procedures developed within GRUAN (Section 10), and then submitted by the Lead Centre to the GRUAN data repository (currently NOAA's National Climatic Data Center, NCDC; Section 8.6).

The GCOS Secretariat provides additional support to the GCOS Steering Committee, the AOPC, the WG-GRUAN and the GRUAN Lead Centre.

GRUAN sites shall use a designated system of methods, techniques and facilities, implemented for making and archiving best quality upper-air observations on a global scale. At any site, this system will not be changed without prior notice to the GRUAN Lead Centre. GRUAN operations shall integrate where possible and when feasible with other international climate monitoring programmes. GRUAN operations will incorporate an assurance programme to validate the stability of the measurement series and uncertainty of the measurements, agreed with WG-GRUAN, and managed in detail by the Lead Centre.

GRUAN shall also be responsive to the latest technological and scientific progress in measurement techniques and observational requirements. Development work can continue at a site until mature and validated, after which it may be introduced into GRUAN operations with the agreement of the Lead Centre. This process will follow the GRUAN Management of Change procedures (see Section 2.3).

The design of GRUAN shall recognise the heterogeneity of the network of sites, many of which will have primary responsibility to networks other than GRUAN.

1.4. Implementation of GRUAN

The implementation of GRUAN shall be guided by the WG-GRUAN. Specific issues to be investigated in support of GRUAN implementation shall be performed either by the WG-GRUAN or by GRUAN task teams established by the WG-GRUAN. These task teams will entrain operational and other relevant expertise in support of GRUAN and will work in coordination with the GRUAN Lead Centre.

A GRUAN Analysis Team for Network Design and Operations Research (GATNDOR) shall undertake focused, short-term research to address specific topics identified by the WG-GRUAN. The work will be conducted in coordination with other relevant GRUAN task teams. GATNDOR activities shall be coordinated with the GRUAN task teams and with national GCOS programmes when appropriate.

The WG-GRUAN shall agree on the appropriate method of establishing standard operational procedures for all observing systems within GRUAN. This could be a new task team, including investigations at the Lead Centre, or an existing instrument team within other associated WMO projects/operational groups. The task teams shall evaluate the appropriateness of uncertainty estimates, the usefulness of particular measurements and operational procedures, synthesize the available knowledge, and develop recommendations to improve GRUAN measurements and operations. These task teams shall confer regularly to evaluate the current status of GRUAN observations, to identify weaknesses, and to incorporate new scientific understanding into GRUAN. The expertise of these teams shall also be used to support the Lead Centre in guiding individual sites through changes in instrumentation and operating procedures without impacting long-term measurement time series.

The GRUAN Lead Centre shall identify sites where instrument operators need training, and organise cost-efficient training courses for the network at appropriate locations, as advised by the appropriate task teams, to encourage uniformity of instrument operation between sites. The Lead Centre may liaise with National Metrological Institutes in this regard.

All activities associated with the implementation of GRUAN are the responsibility of the country hosting the GRUAN site and should, as far as possible, be met through national funding.

To best serve the needs of the climate monitoring and research communities, it is essential that GRUAN is cognizant of the evolving science that drives the measurements and accuracy of the GRUAN data. The instrumentation deployed and the observing schedules may differ between sites, as agreed with WG-GRUAN as part of the site assessment and certification process, but the methods of observation used with the main observing systems are expected to be uniform between all GRUAN sites.

1.5. Links to partner networks

The purpose of this section is to provide, as early as possible in this Guide, a context for GRUAN in the broader community of climate and meteorological monitoring networks. For instance, in the original charter for GRUAN (GCOS-92) it is stated that 'where feasible, these reference sites should be co-located and consolidated with other climate monitoring instrumentation'.

GRUAN shall not operate in isolation of existing networks and GRUAN is not intended to replace existing networks in any way. Many initial and candidate GRUAN sites already belong to existing networks such as GUAN, GAW, NDACC, BSRN and SHADOZ. One of the essential characteristics of a successful GRUAN shall be close coordination with the user community and many of these networks are also likely to be users of GRUAN data. Similarly, complementary measurements from these other networks should be collated in a database to enable cross-comparison and to quantitatively link GRUAN measurements to similar measurements made within other networks. As a result, close coordination between the governing bodies of these networks and with the WG-GRUAN is required on a continuous basis. This close coordination can be achieved by having members of the WG-GRUAN attend steering group meetings of partner networks and by inviting co-chairs or steering group members from partner networks and projects to attend WG-GRUAN meetings.

There is a wide range of tools and methodologies that have been developed in existing networks that GRUAN can adopt, extend if necessary, and learn from. Similarly, existing networks will have skills and expertise likely to be useful to GRUAN and its operations. As a result, contact with expert teams from existing networks shall be made by the Lead Centre, WG-GRUAN, GRUAN task teams, and GATNDOR to support GRUAN operations and to avoid duplication of effort by utilizing existing scientific knowledge.

A number of networks currently in operation make measurements which fall within the scope of GRUAN e.g. those sites that make upper-air measurements that are not part of the typical meteorological measurements of temperature, pressure and water vapour. Many of these networks, including GUAN, have developed systems for assuring the quality of their measurements. Where the systems currently in place are sufficient to meet the operational requirements of GRUAN, they should be used by GRUAN. Where networks are working towards QA/QC procedures, GRUAN should partner with these networks to develop systems that meet the operational requirements of both parties. In some cases, sites within these partner networks may also become GRUAN sites. This is encouraged since it facilitates a traceable link between GRUAN measurements and measurements made at all other sites within the partner network (assuming that the measurements within the partner network are traceable and can be quantitatively linked).

Existing networks and potential resources from within those networks likely to be of value to GRUAN are discussed below.

1.5.1. GOS (Global Observing System)

The GOS provides a coordinated system of methods and facilities for making meteorological and other environmental observations on a global scale in support of all WMO Programmes; the system is comprised of operationally reliable surface-based and space-based subsystems. The GOS comprises observing facilities on land, at sea, in the air and in outer space. These facilities are owned and operated by the Member countries of WMO each of which undertakes to meet certain responsibilities in the agreed global scheme so that all countries can benefit from the consolidated efforts. Since GRUAN will comprise an important component of the GOS, it is necessary that GRUAN operations are cognizant of, and engage with, all related components of the GOS.

1.5.2. GUAN

As noted above, GRUAN will provide a reference backbone for GUAN. The greater the number of GUAN sites that become GRUAN sites, the more efficiently the outcomes of GRUAN will transfer to GUAN. Where GRUAN sites are operating as NMHS (National Meteorological and Hydrometeorological Services) sites, new measurement methodologies developed at those GRUAN sites should efficiently propagate to other GUAN sites operated by the same NMHS.

1.5.3. GAW (Global Atmosphere Watch)

The GAW programme of WMO is a partnership involving 80 countries, providing reliable scientific data and information on the chemical composition of the atmosphere, and the natural and anthropogenic drivers of changes in chemical composition. In this way, GAW improves understanding of the interactions between the atmosphere, the oceans and the biosphere. GAW has strong links to GCOS and so is likely to have skills and resources that could be used to support GRUAN.

1.5.4. NDACC (Network for the Detection of Atmospheric Composition Change)

NDACC, which reports to GAW, comprises more than 70 remote-sensing research sites for observing and understanding the physical and chemical state of the stratosphere and upper troposphere and for assessing the impact of stratospheric changes on the underlying troposphere and on global climate. NDACC currently incorporates 5 water vapour measurement sites and up to 8 temperature measurement sites. There are a number of key differences between NDACC and GRUAN that require GRUAN to operate as a new and independent network, including:

- NDACC aims to observe and understand the chemical composition of the stratosphere and upper troposphere. For GRUAN the highest priority observations are the atmospheric state variables of temperature, pressure and water vapour.
- The primary focus of NDACC is on ozone and the chemicals responsible for ozone depletion. The primary focus of GRUAN is on climate and the processes driving changes in climate.
- NDACC operates as a federation of independent measurement sites. NDACC does have in place stringent standards which must be met for measurement programmes to become part of the network. Conversely, GRUAN requires coordination by a Lead Centre that implements a minimum set of standard operating procedures across the network as a whole to ensure cross-site comparability of instrument suites.

There are, however, a number of measurements and operational procedures common to both networks and every effort should be made to avoid duplication of effort and to ensure that the lessons learned within NDACC are assimilated into GRUAN. For example:

- NDACC has established 'working groups', many of which focus on specific instruments used within NDACC. GRUAN task teams currently include a mix of teams focussing on specific measurements systems (radiosondes and precipitable water from GNSS) and on network wide operational issues. As more measurement systems are incorporated into GRUAN operations, consideration should be given to later expanding the ‘Ancillary Measurements’ Task Team to include specific measurement systems in addition to the 'cross-cutting' task teams that focus on issues common to the network as a whole.

- Measurements of vertical ozone and water vapour profiles made within NDACC will be common to measurements made within GRUAN. This includes both balloon-sonde and lidar measurements.
- Techniques have been developed within NDACC to manage changes in instrumentation. GRUAN should build off the expertise developed in this community over the past two decades e.g.
    i) The JOSIE ozonesonde intercomparisons (Smit et al., 2007).
    ii) Regional ozone profile intercomparisons from multiple instruments (McDermid et al., 1998a; McDermid et al., 1998b).
    iii) Intercomparisons of vertical water vapour profile measurements (Leblanc et al., 2011; Hurst et al., 2011a).
- Measurement redundancy in the NDACC sites has been a strength of the network since it allows intercomparisons of supposedly identical measurements by different instruments which often highlight previously unknown deficiencies in the measurements (Brinksma et al., 2000). GRUAN will include similar measurement redundancy (see Section 6.2).

1.5.5. BSRN (Baseline Surface Radiation Network)

BSRN provides a worldwide network to continuously measure radiative fluxes at the Earth's surface. The network comprises about 40 sites between 80°N and 90°S many of which began operation in 1992 and each year more sites are added to the network. These sites provide data for the calibration of measurements made within the GEWEX Surface Radiation Budget (SRB) Project and other satellite-based measurements of radiative fluxes. BSRN data, archived at the Alfred Wegener Institute (AWI) in Bremerhaven, Germany, are also used to validate radiative flux models. In 2004, BSRN was designated as the global surface radiation network for GCOS. The BSRN sites also contribute to GAW (see Section 1.5.3).

The primary goal of BSRN is to monitor the shortwave and longwave radiative components of the Earth radiation budget and their changes with the best methods currently available. Therefore the measurements of longwave and shortwave downward and upward radiation within GRUAN will overlap with the measurements made within BSRN. Access to the BSRN calibration facilities at the Physikalisch-Meteorologisches Observatorium Davos (PMOD)/World Radiation Centre (WRC) is highly advantageous to GRUAN. BSRN includes a working group on measurement uncertainties that can provide guidance for establishing the radiation measurement uncertainties within GRUAN.

1.5.6. WOUDC (World Ozone and UV Data Centre)

- The WOUDC is one of the World Data Centres which are part of the GAW (see Section 1.5.3) programme of WMO. The WOUDC, operated by the Experimental Studies Section of Environment Canada in Toronto, is not so much a network as an international repository for ozone and UV data. There are many practices employed within the ozone measurement community that are likely to be useful to GRUAN. For example, the management of the Dobson Spectrophotometer and Brewer Spectroradiometer networks, both of which provide data to the WOUDC, demonstrate many of the principles that form the foundation for GRUAN. These include: Undertaking regular regional intercomparisons of instruments which always include a travelling standard (such as the primary reference Dobson instrument) which facilitates standardization of instrument performance between regions.

- Archiving of raw data to permit later reprocessing should new improved ancillary data become available e.g. the shift to the Bass and Paur ozone absorption cross-sections in the late 1980s. A similar process is now underway to evaluate a possible change from the Bass and Paur cross-sections to e.g. the Daumont (Daumont et al., 1992) cross-sections.
- Careful QA/QC of data before archiving and strict version control of data submitted to international archives.

These principles have resulted in ground-based total column ozone time series of sufficient quality to allow detection of the multi-decade decline in ozone until the end of the 20th century and the onset of ozone increases thereafter.

1.5.7. SHADOZ (Southern Hemisphere Additional Ozonesondes)

The SHADOZ project was initiated to remedy the lack of consistent tropical ozonesonde observations. This was done by increasing the frequency, and improving the quality, of ozonesonde launches at selected tropical ozone observing sites (Thompson et al., 2003). Rather than establishing an entirely new network, SHADOZ aims to enhance ozonesonde launches at existing facilities on a cost-share basis with international partners. The geographical coverage of the network was specifically designed to address targeted research questions.

1.5.8. AERONET

AERONET (AErosol RObotic NETwork) is a federation of ground-based remote sensing aerosol networks with contributions from national agencies, institutes, universities, individual scientists, and research partners. The programme provides a long-term, continuous and publically accessible database of aerosol optical, microphysical and radiative properties. The standardization of instruments, calibration procedures, and data processing and distribution is well aligned with the needs of GRUAN.

The AERONET programme provides globally distributed observations of aerosol optical depth (AOD) at different wavelengths, products derived from the raw measurements, and precipitable water in diverse aerosol regimes. Aerosol optical depth data are computed for three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened), and Level 2.0 (cloud-screened and quality-assured). It is primarily the level 2.0 data that are likely to be of interest to GRUAN since these data are quality-assured. Inversions, precipitable water, and other AOD-dependent products are derived from these levels and may necessitate additional quality checks.

1.5.9. European Aerosol Research Lidar Network (EARLINET)

EARLINET, which was established in 2000, is the first aerosol lidar network whose main goal is to provide a comprehensive, quantitative database of the three-dimensional time evolving distribution of aerosols over Europe. The network currently comprises 28 sites broadly distributed over Europe. Network-wide observations take place on a fixed schedule. To collect temporally unbiased data, all sites perform measurements on two fixed days each week. Lidar surveillance adheres to a regular timetable of once each week around noon when the boundary layer is usually well developed, and twice weekly at night under low background light conditions in order to perform Raman extinction measurements. Other network activities monitor extraordinary events such as Saharan dust outbreaks, forest fires, photochemical smog, and volcanic eruptions.

A rigorous quality assurance programme is applied to both instruments and evaluation algorithms, and a standardized data exchange format has been implemented. EARLINET has defined a standard set of primary aerosol optical parameters including profiles of aerosol extinction and backscatter coefficient, lidar ratio, and the Ångström exponent. Standardization is also implemented at instrument and algorithm levels; standardized processing developed in the EARLINET ASOS project, provides all EARLINET partners with a method to analyse their data from raw lidar signals to final products in an automated and standardized way.

EARLINET represents the best available tool for validation and exploitation of data from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) mission, and future new generation satellite missions with on-board lidars. In cooperation with a number of networks distributed worldwide, EARLINET is also making a significant contribution to implementing the GAW Atmospheric Lidar Observation Network (GALION). EARLINET can be expected to contribute to defining and standardizing the aerosol related products within GRUAN.

1.5.10. Atmospheric Radiation Measurement (ARM) Programme

The goal of the U.S. Department of Energy ARM programme is to study alterations in climate, land productivity, oceans or other water resources, atmospheric chemistry, and ecological systems that may alter the capacity of the Earth to sustain life. This includes improving the atmospheric data sets used in regional and global climate models. A primary objective of the ARM user facility is improved scientific understanding of the fundamental physics related to interactions between clouds and radiative feedback processes in the atmosphere.

Five of the current GRUAN sites are also ARM sites, in part because the radiation measurements made at these sites satisfy many of the ECV measurement requirements within GRUAN. The dedicated Data Quality (DQ) Office which ARM established in July 2000 to coordinate and implement efforts to ensure the quality of the data collected by ARM field instrumentation will likely provide a number of tools which could be implemented across the GRUAN network to ensure the quality and network homogeneity of the radiation measurements. The DQ Office is responsible for ensuring that quality controlled results are communicated to data users so that they may make informed decisions when using the data, and to ARM's Site Operators and Engineers to facilitate improved instrument performance and thereby minimize the amount of unacceptable data collected. The ARM DQ Office has developed a suite of sophisticated data quality visualisation tools that are likely to be of interest to GRUAN.

Another ARM organizational structure that is likely to be relevant for GRUAN is the assignment of instrument mentors (as recommended in GCOS-112). Because GRUAN task teams are not structured by instrument (as is the case for NDACC where each working group focuses on one instrument), having ARM-type instrument mentors that advise on instrument operation, maintenance and calibration across the network as a whole may be beneficial. Instrument mentors have an excellent understanding of *in situ* and remote-sensing instrumentation theory and operation and have comprehensive knowledge of the scientific questions being addressed with the measurements made. They also possess the technical and analytical skills to develop new data retrievals that provide innovative approaches for creating research-quality data sets.

1.5.11. Partnership with Meteorological agencies

Meteorological agencies producing global real-time analyses (e.g. UKMO, DWD, ECMWF, NCEP, NOAA, JMA) or historical reanalyses (e.g. NCEP/NCAR, NCEP/DOE, NCEP-CFSR, ECMWF, JMA, NOAA and NASA) are likely to be users of the high quality data produced by GRUAN. Well-developed systems exist for monitoring the quality of operational observations, whether it is the performance of individual radiosonde sites or the bias corrections required by current satellite observations. Therefore, diagnostics obtained from the various assimilation schemes used in such activities will provide valuable metadata on the consistency of the GRUAN measurements with other data used in the operational analyses (thereby facilitating easier comparisons of GRUAN measurements with e.g. satellite-based measurements) and on the representativeness of the uncertainty estimates on the GRUAN data. If the GRUAN data are to be used in the 4D-Var assimilation schemes either for ingest or validation purposes, it is essential that the precise 4D (latitude, longitude, altitude and time) coordinates associated with any measurement are available (see Section 8.3). Reference sites will prove essential for helping to characterize observational biases and the impact of observing system changes, as well as to understand model errors, all of which are important aspects in creating high-quality reanalyses (Schubert et al., 2006). Studies that demonstrate the value that GRUAN measurements will add to NWP and to meteorological reanalyses are currently lacking.

Some GRUAN sites may also be National Meteorological Service (NMS) sites, or may be paired with an NMS site to extend the range of measurements performed, with the result that NMSs are likely to provide partial or full support for a site.

1.6. Link to satellite-based measurement programmes

If GRUAN data are traceable to reference measurements they will be useful to the satellite measurement community for validating satellite-based retrieval products and for validating associated radiative transfer (RT) models. GRUAN data sets are also useful for supporting calibration and validation of satellite-based sensors, for example, compensation for offsets and drifts when creating merged inter-satellite products, but this can require access to accurate RT models. The use of GRUAN profile measurements to provide target spectra for RT model validation is key and feeds back to the sensor and product monitoring. Because GRUAN measurements are likely to serve a wide range of end-users within the satellite measurement community, WG-GRUAN members shall be assigned to liaise with key clients within the satellite community (e.g. GSICS, SCOPE-CM), and with other data providers (e.g. the Radiation Panel within GEWEX), to ensure that GRUAN data products are tailored, where possible, to best meet the needs of this community. Once GRUAN datasets become available, routine monitoring and pilot studies of enhanced datasets using these reference measurements need to be undertaken. Measurement scheduling within GRUAN to accommodate satellite validation activities is discussed further in Section 7.4.2.

1.6.1. Validating satellite-based retrievals

Satellite-based retrieval is used to derive atmospheric vertical (or slant) profiles of temperature and water vapour of primary interest to monitor climate. A variety of retrieval approaches including optimal estimation (e.g. Rodgers 2000), using a RT model, and neural networks (Blackwell and Chen 2009) are available. The consistent validation of retrieved products across the landscape of satellites, sensors and scientific algorithms, is challenging and that challenge is exacerbated by the degree of undocumented errors and uncertainties inherent in conventional ground-truth observations. A programme of routine validation of satellite derived products using independent reference profiles measured at GRUAN sites can significantly impact inter-satellite product characterization while facilitating transfer standard feedback to conventional global networks. For cases of candidate satellite-based references, for example, Global Positioning System Radio Occultation (GPS-RO) measurements for upper tropospheric and stratospheric temperature, transfer standard feedback to GRUAN can occur.

Ground-based reference profile measurements may also provide independent standards for special case studies against which the satellite retrievals may be validated. For example, Vömel et al. (2007a) demonstrate how reference-quality *in situ* water vapour measurements available from GRUAN can be used to validate satellite-based observations of stratospheric water vapour.

The extension of the NOAA Product Validation System (NPROVS) (Reale et. al 2012), operated at the NOAA Center for Satellite Applications and Research (STAR), to provide GRUAN products monitoring and integration into operational satellite products validation (and RT model validation, next section) is in development.

1.6.2. Validating forward models

Satellite-based retrieval of atmospheric parameters often relies on an optimal estimation approach (e.g. Rodgers 2000) to derive profile and column density information of these parameters. Optimal estimation employs a forward RT model that is used to simulate the radiance field that a satellite-based sensor would sample for a given state of the atmosphere. These are simulated from an *a priori* or background profile. To retrieve a state vector (the true values of the atmospheric parameters of interest), the simulated values together with the observed and associated uncertainties are inverted. This requires that the observed be adjusted to the RT model to avoid biased retrievals. GRUAN measurements and profiles are uniquely qualified to provide target and calculated spectra, respectively, for use in both RT model validation and the generation of RT bias adjustment coefficients for retrievals. Even with accurate RT, a retrieval inversion remains poorly constrained, i.e. does not have a unique solution, and, as a result, the *a priori* (background) information has a strong influence on the retrieval. While the direct use of GRUAN measurements to provide *a priori* information for operational global retrieval is not possible, indirect use of GRUAN traceable measurements of atmospheric state variables such as temperature, pressure and water vapour that partially define the radiative transfer properties of the atmosphere are uniquely qualified for RT model validation and associated satellite sensor bias adjustment for retrieval (and similarly for NWP assimilation of satellite radiance).

1.6.3. Calibration and validation of satellite-based sensors

GRUAN measurements and products can be used to support the calibration and validation of satellite based sensors. Ohring et al. (2005) show how calibration against ground-truth can unambiguously remove residual biases in observed sensor measurements in order to be useful for climate monitoring. GRUAN and GSICS (Global Space-based Inter-Calibration System) can be complementary in meeting this need.

The need for inter-site homogeneity within GRUAN has special significance for validation of satellite-based measurements. Ground-based measurements made at all GRUAN sites shall be made in a similar fashion so that differences in the soundings of ECVs between GRUAN sites are as small as possible. If this is achieved, differences between radiances calculated from GRUAN measurements and the radiances measured from the satellite will then primarily be a function of systematic bias in the satellite radiance or caused by a difference in other conditions, e.g. thin clouds in the satellite field of view, surface emissivity, etc.

The issue of measurement scheduling within GRUAN to accommodate satellite validation activities is discussed further in Section 7.2.

1.6.4. Creating global homogeneous atmospheric climate data records

While satellite-based observations (measurements and products) have the advantage of providing global or near-global geographical coverage, the quality and usefulness of the data is compromised by an inability to conduct scientifically unambiguous regular calibrations (unless, for example, adequate, SI traceable on-board calibration sources are available), limited vertical resolution and difficulties in continuity due to drifting orbits which, for variables showing strong diurnal variation, can alias into apparent trends. Furthermore, limited satellite sensor lifetimes often require data series from multiple satellites to be spliced together to form long-term data records.

Discontinuities between satellite-based measurements of climate variables can be ruinous for detecting variability and long-term changes in climate. Although it is recognized that GSICS addresses many of these concerns, the reference measurements that GRUAN will produce can help to minimize offsets and drifts between these separate satellite-based measurement series within the limitations imposed by the uncertainties on the GRUAN measurements and in conjunction with RT model validation account for absolute accuracy not directly available in GSICS space-based splicing. In this way GRUAN can provide a true (accurate) reference-measurement to serve as a common baseline when splicing satellite-based measurement time series.

There are many algorithms, based on a large body of existing literature, which can be used to analyse differences between a given satellite-based data set and the GRUAN reference-measurement and then automatically detect steps and drifts in the differences. The underlying systematic structure in such differences can then be used to homogenize the satellite-based measurements with the GRUAN reference standard and compare results from GSICS.

By contributing to the creation of global homogeneous ECV databases, GRUAN will connect to the WMO SCOPE-CM (Sustained, Co-Ordinated Processing of Environmental Satellite Data for Climate Monitoring) programme. The aim of SCOPE-CM is to establish a network of facilities ensuring continuous and sustained provision of high-quality satellite products related to ECVs, on a global scale, responding to the requirements of GCOS. GRUAN and SCOPE-CM shall collaboratively contribute to Action C10 defined in the GCOS implementation plan (GCOS-92) viz. 'Ensure continuity and over-lap of key satellite sensors ... undertaking reprocessing of all data relevant to climate for inclusion in integrated climate analyses and reanalyses' (Action C8 in the 2010 update of the GCOS implementation plan; GCOS-138).

2. REFERENCE MEASUREMENTS

2.1. Terminology

The following terminology, extracted and adapted from the *Guide to the Expression of Uncertainty in Measurement*[1], is used throughout this Guide to describe the uncertainty components of a reference measurement:

True value: This is a value consistent with the definition of a given particular quantity that would be obtained by a perfect measurement. True values are by nature indeterminate.

Measurement accuracy: Every measurement has imperfections that cause it to differ from the true value. The measurement accuracy describes the closeness of the agreement between the result of a measurement and the value of the measurand.

Measurement uncertainty: A parameter, associated with the result of a measurement, which characterizes the dispersion of the values that could reasonably be attributed to the measurand. Measurement uncertainties may be time dependent.

Measurement error: The result of a measurement minus the true value of the measurand.

Random error: The result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatable conditions. The random error component of any measurement is the result of stochastic variation in quantities that influence that measurement. While random errors cannot be designed out of a system, the random error on the mean of multiple measurements is reduced since, by definition, the expected value for the random error is zero. The term ‘random error’ is preferred over the term ‘precision’ since precision is often used to designate the number of bits or significant digits to which a value is specified.

Random uncertainty: Because it is seldom possible when making measurements of the atmosphere to conduct a large number of measurements of the same measurand under repeatable conditions, the concept of a ‘random error’ is seldom applicable. Therefore the term ‘random uncertainty’ is often preferred as this refers to the distribution of potential values with defined confidence limits.

Systematic error: The mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus the true value of the measurand. It results from systematic biases that do not average to zero as the number of measurements increases. However, if these systematic biases can be identified and quantified, they should be corrected for. The term ‘systematic error’ is preferred over the term ‘accuracy’ since it denotes more clearly that the deviation is systematically in one direction.

Stability: Stability refers to the consistency of random errors and systematic errors with time. Undetected changes in systematic errors induce artificial trends in measurement time series.

Independent measurement: Two measurements are considered independent when no aspect, either directly or indirectly, of one method of measurement involves the other.

Correction lifetime: A corrected result is one where a measurement has been corrected for any systematic error. This correction may depend on an independent measurement from another source and may have a finite ‘lifetime’ in the sense that later reprocessing of the measurement may revise the estimate of the systematic error, requiring a new correction.

2.2. The concept of a reference measurement

As denoted by its title, the primary objective of GRUAN is to provide reference measurements for a range of upper-air climate variables. Reference quality atmospheric observations are based on

[1] http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

key concepts in metrology (measurement science), in particular traceability. Metrological traceability is the process whereby a measurement result, i.e. an estimated value and its associated uncertainty, can be related to a reference through a documented, unbroken chain of calibrations, each of which contributes to the measurement error.

A reference measurement result refers our current best estimate of the value for some atmospheric parameter, as well as a best estimate for the level of confidence that is associated with this value, recognising that future improvements in measurement techniques and/or reprocessing following new knowledge may lead to refinements in that reference value. In most cases it will be the best technology available that will achieve the best estimate of the value for some target atmospheric parameter. Reference measurement methods accommodate the unavoidable sources of uncertainty in the compilation of the net measurement uncertainty while excluding those sources of uncertainty that can be avoided. For example, in the pre-deployment calibration of a sensor, there will be some unavoidable uncertainty in the accepted measurement standard and hence some unavoidable uncertainty in the calibration which must then be included in the net measurement uncertainty. However, contributions to measurement uncertainty from e.g. an improperly documented traceability chain, proprietary methods, appeal to physical principles without experimental verification, or the use of an improper calibration standard must be avoided. Similarly, when the instrument is later deployed, there will be numerous, unavoidable contributions to the total measurement uncertainty from e.g. uncertainty in the input data, data processing constants, the data retrieval algorithm, and in the physical/chemical model of the measurement system used to convert raw measurements into data. However, contributions to measurement uncertainty from the use of ‘black box’ software, undocumented or unvalidated measurement adjustments, or the disregard of known biases must be avoided.

A reference data product can be produced from a single reference measurement, by averaging multiple reference measurements over some time period, or by processing reference measurements from multiple instruments as is done for the creation of a SASBE (Site Atmospheric State Best Estimate; Tobin et al., 2006). This highlights the importance of measurement redundancy (see Section 6.2) in that access to coincident multiple measurements of the same quantity often leads to a more robust estimate of the true value and a better estimate of the uncertainty on that value.

The estimate for the level of confidence on any measurement is expressed as the measurement uncertainty. The measurement uncertainty is a property of the measurement that combines instrumental as well as methodological uncertainties. The measurement uncertainty describes the current best knowledge of instrument performance under the conditions encountered during an observation, it describes the factors impacting a measurement as a result of operational procedures, and it makes all factors that contribute to a measurement traceable. Within GRUAN this uncertainty shall be vertically resolved and each measurement in a profile shall be treated as a single measurement result requiring both the estimate and its uncertainty. To provide the best evaluation for the instrumental uncertainty, a detailed understanding of the instrumentation is required for the conditions under which it is used. Specific requirements that an observation must fulfil to serve as a reference for comparing or validating other systems in the GRUAN context, have been defined in Immler et al. (2010).

A reference measurement result typically arises from a defined measurement procedure that involves standards traceable to national or international standards as maintained at National Metrological Institutes (NMIs). For GRUAN, a reference measurement is one where the uncertainty of the calibration and the measurement itself is carefully assessed. This includes the requirement that all known biases have been identified and corrected, and, furthermore, that the uncertainty on these bias corrections has also been determined and reported. An additional requirement for a reference measurement is that the measurement method and associated uncertainties should be accepted by the user community as being appropriate for the application.

The methods by which the measurements are obtained and the data products derived shall be reproducible by any end-user at any time in the future. It should be kept in mind that these end-users are likely to use GRUAN data for decades to come. They shall be able to reproduce how measurements were made, which corrections were applied, and be informed as to what changes occurred during the observation and post-observation periods to the instruments and the algorithms. Hence maintaining comprehensive rich metadata regarding data provenance and processing is key.

> In brief, *reference* within GRUAN means that, at a minimum, the observations are tied to a traceable standard, that the uncertainty on the measurement (including corrections) has been determined, and that the entire measurement procedure and set of processing algorithms are properly documented and accessible.

2.3. Managing change

2.3.1. Guiding principles

GRUAN recognizes that change is inevitable and that changes in:

i) instrumentation,
ii) operating procedures,
iii) data processing algorithms,
iv) instrument operators,
v) location of instruments, and
vi) operating environments for instruments,

collectively referred to hereafter as change events, are all likely to introduce sources of operational uncertainty into GRUAN data products. Some of these changes, rather than being instantaneous and introducing a stepwise change in the time series, may be gradual (e.g. urbanization of the surrounding area or growth of nearby vegetation) and induce a trend-like drift in the measurements. GRUAN appreciates that without change, improvement is impossible. While the primary goal is to avoid unnecessary changes, i.e. those changes that have no scientific, financial or operational benefit, where changes are beneficial and/or necessary, the goal is to manage such changes in a way that the intercomparability of the climate record is maintained across the transition and that the change does not compromise the integrity of the long-term climate record.

The purpose of this section is to describe the protocols for managing change within GRUAN. Items (i) to (iii) above are likely to have network wide impacts while items (iv) to (vi) are site specific and are therefore dealt with separately in Sections 2.3.11 and 2.3.12 respectively.

A goal within the 'Management of Change' research topic of GATNDOR is to provide scientific bases to develop operational practices to better manage the changes listed in items (i) to (iii) at GRUAN sites, and to accurately merge disparate data segments to create a homogeneous time series. Protocols developed by GATNDOR and others, as detailed in this section, are then implemented throughout the network under the mandate of the Lead Centre.

In addition to the following GCOS climate monitoring principles[2] relevant to management of change:

[2] http://www.wmo.int/pages/prog/gcos/documents/GCOS_Climate_Monitoring_Principles.pdf

1. The impact of new systems or changes to existing systems should be assessed prior to implementation.
2. A suitable period of overlap for new and old observing systems is required.

The following are also considered as relevant guiding principles for GRUAN:

3. *Embracing change:* GRUAN must not be resistant to change but must actively encourage carefully managed changes, as required, since this is essential to on-going improvement of the network. However, the advantages of making any change must always be weighed against the inherent disadvantages of making a change.
4. *Change event notification:* A change event begins with measurements being initiated with a new measurement system and ends with the termination of the old measurement system. In GRUAN every change must begin with a change event notification (see Section 2.3.11).
5. *Justification of change:* Any putative change in a measurement system must be fully justified before the change is enacted. The advantages and disadvantages of making the change must be carefully assessed as part of the justification process. Laboratory tests of old and new instruments/sensors, parallel testing of old and new retrieval algorithms, and/or parallel testing of old and new measurement systems (Section 2.3.3) may all be an important part of such an assessment. In GRUAN, justification of change should, in the first instance, fall to the central data processing facility responsible for producing that data product, and any task team specifically dedicated to that product, since they are likely to be best equipped to assess the consequences of that change for data homogeneity across the network as a whole. In addition, the Lead Centre must act as a clearinghouse for all proposed measurement system changes (see Sections 2.3.11 and 2.3.12). Given the wide range of observing systems that potentially may be deployed as part of GRUAN operations, the protocols for (a) assuring high instrument stability and (b) deciding when an improvement merits a change to the GRUAN methods of observations will need to be developed as required by the WG-GRUAN or appropriate task teams, given guidance on the user requirements when required by GATNDOR. For all instruments used in GRUAN, the extent to which sites are willing and able to adopt the standard procedures recommended by the Lead Centre is a determining factor in the added value that a site brings to the network (see Section 5). Similarly the equipment and methods of observation in daily use shall not be changed, without consultation with the WG-GRUAN, as advised by the Lead Centre. Improvements to performance can be developed at GRUAN sites, but the evidence that the improvement justifies changing the GRUAN measurement protocol must be rigorously assessed before any change to GRUAN observations is considered by WG-GRUAN.
6. *Preparing for change:* A quantitative assessment of the impacts of any planned change must be undertaken before the implementation of the change. The assessment must continue through the change period and must include not only the impact of the change on the measurement, but also the impact on the uncertainty on the measurement. The process of quantifying these impacts will depend on the nature of the change. The impacts of a change in sensor should be quantified through laboratory studies in such a way that knowledge of the new sensor is at least as detailed as knowledge of the old sensor. The impacts of a change in calibration should be quantified through traceability of the calibration standard and method. The impacts of a change in processing algorithm should be quantified by applying the old and new algorithms to a diverse set of common data.
7. *Validating impacts:* If a change has been properly managed through careful preparation, quantitative assessment of the impacts of the change on both the measurement and its uncertainty, and incorporation of that understanding into the processing chain (which may re-

quire reprocessing of historical data – see below), no unexpected discontinuities in the measurement series should result. Validation of the process can be achieved by subjecting the entire measurement series to homogenisation tests. Significant resources and techniques have already been developed within the surface climate community[3] and upper-air climate community to detect inhomogeneities in climate records (e.g. Seidel et al., 2010; Thorne et al., 2010; Dai et al., 2011) although to do so for upper-air records is more challenging than for surface climate records (Thorne et al., 2005). Impacts of changes must be assessed in light of the different intended uses of GRUAN data products, viz.:

i) *Trend detection:* Changes in measurement uncertainties following measurement system changes will affect the statistical significance of a derived trend in the long-term data record (see e.g. Stolarski and Frith, 2006; Seidel and Free, 2006).

ii) *Satellite calibration/validation:* While satellite calibration/validation should not be impacted by a managed transition from an old to a new measurement system within GRUAN, such transitions should be avoided during any planned intensive satellite calibration/validation campaign or known satellite platform transition, to the extent possible.

iii) *Process studies:* For studies where insight is gained from analysis of long-term time series, ensuring a homogeneous data record will remain a priority.

iv) *Input to NWP and meteorological reanalyses:* Long-term stability of NWP systems require long-term homogeneity of the observations used as input. As is clear from the discontinuities in the stratospheric temperature record in reanalyses, ensuring long-term homogeneity of the records ingested in reanalyses is critical for ensuring the quality of the reanalyses.

8. *Change and uncertainty:* Knowledge of any measurement system can never be complete or perfect. Transitioning from an old to a new measurement system always introduces a new source of uncertainty which must be captured in the uncertainty estimate on the measurements. While every effort must be made to ensure that the change is properly managed such that systematic biases and/or drifts between the old and new instrument systems is minimized, it must be recognized that any measurement change will also change the uncertainty on the measurement series.

9. *Network homogeneity:* Managing change is essential to maintaining network homogeneity. If changes are implemented unilaterally at a single site, and even if those changes are implemented such that the long-term homogeneity of the measurement record at that site is preserved, the change may introduce inconsistency with other sites in the network. Changes in measurement systems at GRUAN sites should therefore be conducted in such a way that the homogeneity of the resultant GRUAN data products across the network is not compromised. This does not necessarily mean, for example, that any change in instrumentation must be implemented at all sites at the same time (which may be detrimental to the management of that change) but rather that change at any one site must be conducted within the context of, and in consultation with, other sites in the network. The Lead Centre shall play a key role in ensuring such smooth transitions.

10. *Supporting reprocessing:* As new and more in-depth knowledge of various measurement systems is gained, and in particular following change events, reprocessing of historical data may be necessary. Such reprocessing will require revision of the homogenization procedures applied at each previous change event to produce a homogenised data record. It is essential, therefore, that raw data, as was well as detailed metadata collected during change

[3] see e.g. http://www.homogenization.org

events, are archived so that such reprocessing can be easily achieved. This is discussed in greater detail in Section 2.3.4.

11. *Single changes:* Whenever a measurement system is changed, as many similarities as possible between the old and new systems should be maintained e.g. both the instrument and its location should not be simultaneously changed. Multiple simultaneous changes must be avoided so that the quantitative assessment of the impact of the change on the measurement and its uncertainty is not confounded with other, simultaneous, assessments.

12. *Monitoring changes:* Most changes are planned and therefore can be managed. However, some changes may be unplanned and occur sufficiently slowly that they are not immediately identified e.g. a slow drift in the response of an instrument. Constant vigilance to proactively detect and correct for such changes is required. This can be achieved, in part, through comparison with independent redundant measurements (see 13 below), models (see 14 below), or meteorological reanalyses, and scheduling calibration frequencies according to instrument use.

13. *Use of independent, redundant measurements:* Redundancy in measurement systems provides a powerful tool for validating the management of changes in any one of those systems. Tests of the intercomparability of the system undergoing change with other measurement systems in the set before and after the change can validate the robustness of the management of the change. If the change has been managed correctly, no unexpected differences between the system undergoing change and the redundant measurements systems should be detectable. To take advantage of measurement system redundancy in this way, it is essential that these independent systems are not changed simultaneously.

14. *Use of models:* Where changes in an historical measurement record have not been adequately managed, and where physical or statistical models can faithfully reproduce the key characteristics of the measurement record, the model time series can provide a means of detecting and correcting for systematic biases between old and new measurement systems. For example, comparison of radiation measurements on cloudless days with output from a clear-sky radiative transfer model (Bodeker and McKenzie, 1996) has been used to identify and correct offsets and drifts in surface radiation measurements resulting from changes in instruments or instrument calibration. Statistical models may be of the form of regression models that are fitted to measurements from the existing system and then projected forward to cover the period sampled by the new system, or could rely on measurements from surrounding sites to estimate values at the site of interest. In GRUAN, where all changes are managed changes, the use of models for this purpose should not be necessary.

15. *Instrument calibration:* When instruments are calibrated and traceable to standards, changes in instrumentation can be more easily managed.

16. *Manufacturer involvement:* Efforts must be undertaken to avoid unknown changes e.g. the instrument manufacturer making unannounced changes. GRUAN needs to establish close working relationships with instrument manufacturers so that any changes implemented in the manufacturing of an instrument are made know to the GRUAN community, preferably substantially in advance of deployment, to permit sufficient time to investigate, understand and document the change and its likely impacts.

2.3.2. The importance of metadata

Seldom are metadata more important than when documenting network changes. Complete metadata should include a full account of the operation of the site from as early as possible but at the very least from the date of its certification as a GRUAN site to the present (see Section 8.3).

Detailed archiving of instrument metadata will be vital to managing changes in instrumentation. This will allow later reprocessing of the raw data as 'deeply' as possible (see Section 2.3.4). Since it is not always known in advance which metadata are likely to be required for reprocessing at a later date, GRUAN operators should err on the side of collating as much metadata as possible about measurement systems even if no immediate use for those data can be envisaged. In all cases sufficient metadata must be available to tie the new instrument via a comparable traceability chain to recognized standards.

Metadata should include, for example, geo-tagged and time stamped digital images of the instrumentation used, manufacturer, model, year and serial number of the instrument, key steps in instrument calibration, key steps in the measurement process, the measurement site and surrounding region. Pictures may capture information not initially considered to be relevant but later found to be useful in assessing causes of changes.

A detailed description of how each change in a measurement system was managed is an essential component of the instrument metadata. These metadata must include everything related to the quantitative assessment of the impact of the change on the measurement and its uncertainty. It is particularly important that these metadata identify any sources of uncertainty that could not be quantified when making the change. Access to these metadata associated with change management will be essential for any required reprocessing of the historical record. As much of the process metadata as possible should be directly tied to the data through header field tags unique to each observation. Formats such as NetCDF enable such metadata discovery.

2.3.3. Validating managed changes using parallel observations

Applicability: As detailed in the GCOS Climate Monitoring Principles, parallel operation of old and new measurement systems for an overlap period prior to decommissioning the original system is considered to be a valuable option for managing change. However, within GRUAN, where instruments are calibrated to traceable and fundamental standards, when an old instrument is replaced with a new instrument that is calibrated to the same standards, any discontinuities between the two systems can be determined quantitatively in the laboratory on the basis of their calibrations. In this case parallel observations are not formally required to derive a homogeneous measurement and measurement uncertainty time series, as long as the laboratory assessment covers the range of potential external influences that may affect the measurements. Parallel observations in the field using the new and old systems provide a powerful validation of the laboratory-based results under the actual conditions of measurement. Such redundancy in quantifying inter-instrument differences is encouraged within GRUAN. The objective is to retain the original measurement system and to establish the new system in a manner that maintains as much as possible of the old system: same location, procedures, and sensors; and to document in the associated metadata those elements of the new system that have changed.

Overlap regimen design: As detailed in the second of the GCOS climate monitoring principles, it is essential to ensure that when transitioning from older to newer instrumentation, that a sample of coincident measurements, sufficient to validate *a priori* laboratory-based determination of any biases between the two systems (in the form of a transfer function), is obtained before the evidence is presented to the appropriate GRUAN task team. The length of time for which the old and new systems should be run in parallel, and the frequency with which coincident measurements should be made, will depend on the instruments used, an in-depth understanding of the measurement technique, and the main applications for the long-term measurement record. This may require, for example, more than one parallel observations period e.g. after a nominal initial 6 month overlap period, it could prove valuable to conduct a second parallel testing phase 2 years later to gauge whether there has been any drift in the bias between the old and new systems. In all cases, sound scientific bases should be established to determine the period and frequency of parallel observations. For example, when a change in radiosonde type used at one site is proposed, the old and new sondes should be launched on the same balloon, or if that is not possible (though this would be the exception), as close in time as possible on consecutive balloons, for a period sufficiently long to capture their systematic differences. Analysis of dual sonde data from Lindenberg indicates that about 200 dual sonde flights, sampling both daytime and night-time conditions, over a period of one year is required to achieve 0.05K and 0.3% comparability for temperature and relative humidity, respectively, and to accurately assess the bias between old and new sondes. The number of dual sondes required may be site dependent and will therefore require site specific analysis to determine the number of dual flights and the length of overlap period.

Operational constraints: From an operational perspective, finances and other operational considerations (e.g. availability of staff, land, and the feasibility of maintaining the operation of the original measurement system) will often be limiting factors in defining the duration of the period of parallel observations. Because of the extra demands that such parallel observations place on already stretched financial and human resources, parallel observations should be continued for no longer than required and should be informed by the initial quantitative assessment of the impact of any planned change. For some measurement systems, adequate sampling of the diurnal cycle may also be necessary. The costs associated with running the two systems in parallel should be included in the budget for implementing the change.

Site specific considerations: The overlap period may also depend on the site since seasonal variability may differ between sites such that a site experiencing greater atmospheric variability may take longer to derive a robust estimate of differences between measurement systems than a site experiencing lower atmospheric variability.

Use of statistical analyses: When it is not feasible to operate older and newer measurement systems side-by-side for extended periods e.g. with balloon-borne instruments, alternating between the newer and older systems can also provide a means of validating laboratory-derived quantities used in the management of the change. Various statistical techniques (e.g. regression analysis) can be used to determine whether any systematic differences between the two measurement sets remain after a managed change. These biases can be derived as functions of other variables such as air pressure, temperature, time of day, solar zenith angle etc.

Use of redundant, independent measurements: When parallel observations of old and new measurement systems are not feasible, the availability of additional redundant systems, measuring the same variable with similar sampling attributes (vertical resolution, temporal sampling frequency etc.) is essential for validating a managed change. In such cases an evaluation of the period of overlap of the redundant system(s) with the old and new system, required to validate the robustness of the change management, must be undertaken. When using redundant systems in this way the overlap period will be informed by the initial quantitative assessment of the impact of the change.

2.3.4. Data reprocessing

Reprocessing triggers: Protocols must be established by the designated central processing centre for each GRUAN data product to indicate when reprocessing of the full measurement record at any site is justified or required. Since there is a time and administrative cost associated with the reprocessing of a record, such reprocessing should only be undertaken when justified. This is likely to occur

i) After each change event, and
ii) As new and more in-depth knowledge of various measurement systems is acquired.

Selecting the standard: If the newest part of the record is considered to be the reference result, then the entire historical record must be reprocessed to bring it in line with the newest part of the record. It is also possible, however, that the existing record is recognised as the standard in which case the information obtained from the quantitative assessment of the impact of the change is applied to the newest measurements, and their uncertainties, to splice them into the historical record.

Data versioning: Every reprocessing generating a new homogeneous time series over the complete measurement period must be reflected in an increment in the data version as prescribed in the data versioning protocols developed by the Lead Centre. Such data updates must also be communicated to users who have accessed earlier versions of the data and who have voluntarily registered to receive notifications of such data updates (see Section 8.6). For this reason it is also important that all older versions of any data set are always made available through the GRUAN archives.

2.3.5. Managing changes in instrumentation

Triggers for changes in instrumentation: The instruments used in GRUAN are likely to change when:

i) Newer instruments or sensors become available that permit more accurate measurements of the atmospheric state, or more relevant measurements of the atmospheric state.

ii) Cheaper instruments become available that permit higher temporal sampling of the atmosphere at a similar cost to the older system. A cost-benefit analysis, considering all four primary uses of GRUAN data, should be undertaken by the WG-GRUAN, or a body designated by the WG-GRUAN, to inform the decision on whether or not a change in measurement system is justified. Given that uncertainty in trends is often dominated by the contribution of natural variability in the signal, a statistically more robust trend detection may be possible with increased measurement frequency even if the precision of each measurement is somewhat reduced (see also Section 7.3). Process studies may be better served when the processes of interest are sampled more frequently. Satellite calibration/validation may be better served when the number of coincidences with GRUAN measurements is maximized. Reanalyses may be better served by more frequent, less precise measurements than less frequent more precise measurements. Evaluation of each of these cost-benefits must be undertaken when a change is proposed.

iii) The necessities of production engineering. When the components required to build instruments become unavailable or too expensive, or when entire instruments are no longer manufactured or become obsolete, changes in instrumentation will be required and the designated central processing facility for that GRUAN data product will need to decide what level of component change requires additional change testing.

iv) Unplanned changes as a result of a loss of a sensor due to breakage/damage, premature aging, or theft.

Adopting a scientifically robust instrument replacement strategy that maximizes the maintenance of long-term climate records will be important for ensuring the integrity of the GRUAN data products in the face of change.

Assessment of the expected changes: Any change in instrumentation or sensors could potentially lead to discontinuities in the long-term time series and, more importantly, to changes in the characterization of the measurement uncertainties. These changes need to be assessed prior to any change event through careful evaluation in calibration laboratories against traceable reference standards. Technical specifications provided by the manufacturer of the instrument must be verified. In addition, the new instrument should be tested in the field against existing systems under different conditions. All test data should be made available as part of the metadata for the new system. Newer sensors or instruments may have very similar characteristics or may differ significantly in their performance. Changes may be as little as an improved calibration coefficient, or as large as using a completely new technique with completely different calibrations, time constants etc.. The expected impact of a managed change must be assessed and a recommendation should be given as to how to best validate this expected change and how to best address new traits that were not present in the old system. The expected change impact will guide how the change is managed and the level of detail that needs to be documented as part of the metadata.

Instrument intercomparisons: Formal instrument intercomparisons will be essential for developing the in-depth understanding required to manage changes from one instrument to another and for informing decisions on the relative advantages and disadvantages of changing instrumentation. For this reason, participation in formal intercomparisons is expected before the adoption of any instrument within the GRUAN network. Outcomes from such intercomparisons will form an important component of the metadata archived at the GRUAN Lead Centre. Such intercomparisons need not necessarily be organized by GRUAN. WMO and partner networks (e.g. NDACC) often run instrument intercomparison campaigns and GRUAN should participate in these and share the data where possible. Such participation would be mutually beneficial to both communities. GRUAN needs to work closely with CBS and CIMO to gain maximum benefit for all parties from these intercomparisons. In addition to intercomparisons of similar instruments (e.g. radiosondes), intercomparisons between different instruments measuring the same ECV will also be highly informative (e.g. comparisons of ozonesondes, ozone lidars and ozone microwave radiometers at a single site). A number of case studies exist which can be used as examples of how to manage changes in instrumentation. For example the impacts of changes from the Meisei RS2-91 type radiosonde to the Vaisala RS92-SGPJ type GPS sonde at Tateno were quantified by conducting dual sonde flights during four intensive observation periods in December 2009, and in March, June and September/October 2010. Flying dual ozonesondes has proven to be useful when shifting from one ozonesonde system to another or from one standard operating procedure to another (Boyd et al., 1998).

Travelling standards: Travelling standards, or a travelling standard instrument, contribute to the maintenance of network homogeneity when rotated through the sites in the network. Such standards are also essential for validating measurement uncertainties.

Multi-site instrument changes: Consideration will need to be given to the desired strategy when more than one site in the network is making an identical (or very similar) change with respect to timing, sharing of data, and whether certain sites will act as pioneers. This will be especially important where the change is forced by a supply issue. Multi-site instrument changes will require close cooperation between the different sites that will be impacted by the change. Such a multi-site response to instrument changes is also likely to benefit from economies of scale.

Measurement redundancy: Measurement redundancy (see Section 2.3.3 and Section 6.2) has significant benefits for managing instrument change as a second instrument, measuring the same ECV, can be used as a common reference against which both old and new instruments can be compared over an extended period. This benefit increases further when three or more instruments measure the same ECV and any changes are substantially staggered. An ideal aim that assures the record is therefore at least triple redundancy. For *in situ* balloon-borne instruments, consistent ground-check routines between new and old instruments will minimize changes in procedural uncertainty contributions. Measurement redundancy is particularly important in the case of a hiatus between old and new measurement systems e.g. if a measurement system fails and is then later replaced. Since no overlap between the old and new systems is possible, the availability of a third system to act as a transfer standard between the old and new systems is essential. When old and new instruments are both calibrated to the same calibration standard, measurement redundancy is less crucial but is still required in this context as a check that the switch from the old to the new instrument in no way compromises the homogeneity of the measurement series.

Links to instrument manufacturers: Dealing with changes in instrumentation will require GRUAN to establish close two-way links to instrument manufacturers. Inclusion of the Association of Hydro-Meteorological Equipment Industry (HMEI) in discussions of instrument change within GRUAN would be advantageous. A productive point of interaction with the different vendors and manufacturers would be the periodic GRUAN participation in the CIMO multi-sensor field campaigns. Engaging the manufacturers in these field campaigns will assist GRUAN not only in evaluating the different sensors but also as a point of interaction with the vendors apart from the limited HMEI attendance at GRUAN meetings. A close cooperation between GRUAN and instrument suppliers will also help GRUAN to better understand industry capabilities and to better quantify instrumental uncertainties. This cooperation will also help suppliers to better understand GRUAN requirements, and the industry would be able to advise GRUAN of its current and prospective abilities to meet these requirements. For many of the parameters of interest (as instruments of required accuracy do not yet exist), GRUAN aims to further their development in cooperation with instrument manufacturers.

2.3.6. Managing changes in operating procedures

Even if instruments themselves do not change, changes in the operating procedures for an instrument may, if not managed correctly, introduce inhomogeneities in a measurement time series.

For the most part, changes in operating procedures should be dealt with in a fashion similar to changes in instrumentation e.g. after quantitative assessment of the impact of the change, reprocessing of historical data to homogenize the time series followed by redistribution of the full data record with an updated version number.

The expectation is that standard operating procedures for all instrument types within GRUAN will be archived at the Lead Centre and that this body of material will be used to advise sites through transitions in operating procedures. As discussed in Section 5.5 the extent to which a site is prepared to conform to GRUAN standard operating procedures will be one of the criteria used when evaluating the potential inclusion of the site in GRUAN.

2.3.7. Managing changes in data processing algorithms

New knowledge and resultant improvements in reduction of raw data to useful measurements are likely to lead to changes in data processing algorithms. As for changes in operating procedures, such changes in data processing algorithms should be dealt with in a fashion similar to changes in instrumentation.

Every change in data processing algorithm must be reflected in a change in version number of the final data product. Because raw data from various GRUAN sites will be processed at a centralized processing facility (either the Lead Centre or some other GRUAN site or a third part facility with particular expertise in that measurement), changes in data processing algorithms will be implemented uniformly across the network.

To achieve homogeneity across the network it is important that individual sites do not independently implement changes in data processing algorithms for data submitted as GRUAN data even if those changes are well documented and follow the prescriptions listed above. This more central, 'top-down' approach to data processing is different from the more decentralized approach employed in other networks. While such enforced conformity incurs an operational cost, the advantage is that end-users of the GRUAN data products will see data homogeneity not only in time for single sites, but also between sites.

In support of maintaining consistency in the use of data processing algorithms within GRUAN, the Lead Centre will maintain an archive of data processing algorithms which then also comprises an important part of the metadata archive for GRUAN.

Tension may arise where a site may wish to implement a non-standard (at least non-standard for GRUAN) data processing algorithm for some purpose e.g. to create a data product that is tailored for a specific need. Such eventualities can be accommodated by having a central processing facility for each GRUAN product (see above) where a common data processing procedure is applied to the 'rawest' form of data collected. This would not preclude a site from implementing non-standard processing of the raw data and using this for their own purposes.

2.3.8. Managing changes in operators

Ideally the quality of the measurements should be immune from changes in operators. This is more likely achievable if standard operating procedures are developed where there is reduced opportunity for idiosyncrasies of operators to affect the measurements. Metadata should include codes (not names to protect the privacy of operators) to denote where different operators have been responsible for measurements. This also applies to different operators carrying out the data analysis.

2.3.9. Managing changes in instrument location

Even though an instrument may not change, the location of the instrument may change. The instrument may be relocated at a site with a resultant change in operating environment, or may be relocated to a different site. In both cases the 'old' and 'new' system cannot be run side-by-side to establish systematic biases and drifts. Differences between the old and new systems will be some combination of temporal changes in the parameter being measured and changes induced by the horizontal and vertical separation of the old and new instrument locations. In turn, differences resulting from the spatial separation are caused by spatial gradients in the parameter being measured and perhaps also by differences in the operating environment which may induce a non-physical bias between the old and new systems. Two different scenarios must be considered, viz.:

i) *An independent measure of the spatial gradient is available:* This may be available e.g. from satellite-based measurements of the climate variable field. In this case one additional redundant measurement system is required. When relocation of an instrument (system A) is envisaged, that system is operated alongside the redundant system (system B) for a period sufficient to establish any systematic biases or drifts between the two systems. After system A is relocated, simultaneous measurements by systems A and B can be compared after 1) a correction has been made for the effects of the spatial gradient in the climate variable being measured by systems A and B, and 2) a correction has been made for any systematic biases and drifts between the two systems as established during their original period of collocation. Any remaining differences result from changes to the operating environment which can then also be corrected for. It is important that any temporal dependence in the spatial gradient is also captured i.e. it might be necessary to have such a field available at each synoptic time of simultaneous measurement.

ii) *An independent measure of the spatial gradient is not available:* In this case two additional redundant measurement systems are required. When relocation of an instrument (system A) is envisaged, that system is operated alongside two redundant systems (B and C) for a period sufficient to establish any systematic biases or drifts between all three systems. When system A is relocated, so is system C. Differences between systems B and C, after being corrected for their respective biases/drifts, quantify the effects of the spatial separation, while differences between A and C, after being corrected for their respective biases/drifts, quantify the effects of changes in the operating environment (assuming that systems A and C are not similarly affected by changes in the operating environment). Differences between systems A and B test for consistency (closure of the bias budget) between the three systems.

Even with such careful management of location changes, under the protocols developed to evaluate instrument co-location (see Section 6.5), it may be deemed that the new location constitutes a new site within GRUAN and then becomes subject to the GRUAN site assessment and certification process (see Section 5.5).

2.3.10. Managing changes in operating environments

Construction of new buildings or trees being planted or removed at a site may alter the field of view of an instrument. Changes such as the painting of a Stevenson screen may affect temperature measurements. Changes in development around the site may alter the surface albedo of the surrounding area and hence the solar radiation environment sampled by the instrument. It is imperative that all such change events are recorded in the metadata associated with the instrument (log books) and that these events are specifically identified as potential breakpoints in the time series, requiring management, to the central data processing facility. A comprehensive set of photographs providing a horizon-wide view of the site, taken approximately 4 times through the year, and from various reference locations around the site, will provide a valuable resource for assessing changes in the environment at the site. Managing the effects of changes is not simple and is likely to rely on an assessment of the consistency with other data e.g. reanalyses, satellite-based measurements that have been independently verified, or with redundant measurements which are not similarly affected by changes in environmental conditions.

2.3.11. Procedure for multi-site change implementation

In light of the above, and:

i) noting the special importance of change management in GRUAN, and
ii) that sites may not act unilaterally in implementing changes,

the following process for justifying, accepting and implementing changes in:

i) measurement systems,
ii) operating procedures,
iii) data processing algorithms,

shall be followed:

Notification: A change event notification is issued either by the Lead Centre, a GRUAN central processing facility, a GRUAN site, an instrument manufacturer, or another member of the GRUAN community. Proposed changes in operating procedures will likely arise from GRUAN sites, while proposed changes in data processing algorithms will most likely be initiated by the nominated central processing facility for that GRUAN data product. Whatever the origin of the proposed change, the change event notification is documented with the GRUAN Lead Centre.

Assessment: The Lead Centre, in consultation with relevant experts e.g. those at the designated central processing facility for the product affected by the change, makes an initial evaluation of the proposed change. If considered to be worth pursuing, the Lead Centre assesses the advantages, disadvantages, and potential impacts of the proposed change, in particular which parts of the system will most likely be affected. If the knowledge required to quantitatively assess the impact already exists, it is immediately encapsulated in the metadata associated with the change event. If additional studies are required, such studies must be either undertaken by the Lead Centre or commissioned by the Lead Centre. The information and data required to manage the change are captured in a change evaluation report which becomes a key component of the metadata associated with the change.

Consideration will be given as to whether the proposed change should be implemented at a single site or across the network as a whole.

Single site implementation: The change evaluation report, and the timeline for the managed change, will be negotiated with the site and, based on that report, the site will decide on whether or not to implement the proposed change. This timeline includes the actual start of the change, the expected completion date of the change, the expected sequence of dual observations, and the proposed laboratory studies to provide the theoretical basis for the change. While the schedule of simultaneous observations is negotiable, it must be guaranteed that the regular observations schedule is not interrupted. During this time the agreed upon laboratory studies are conducted. The change event ends when the theoretical studies have been completed and when the final report has been written. In cases when laboratory studies cannot be reconciled with the results derived from *in situ* simultaneous observations, this must be noted as part of the metadata, the respective uncertainties increased appropriately, and a proposal developed for how to resolve the discrepancy. If the site decides to proceed and implement the change, any data and metadata collected as part of the change process, as well as a full report on how the change was managed and implemented, must be submitted to the Lead Centre within 3 months of the completion of the switch so that this information can be archived as part of the metadata record for that measurement series from that site.

Network wide implementation: In addition to considering the change evaluation report, the Lead Centre will consult with users of GRUAN data products and, depending on the nature and magnitude of the proposed change, with other climate data bodies such as GCOS, WCDMP and WDAC to thoroughly evaluate the potential implications of network wide implementation of the proposed change.

If it is decided to proceed with network wide implementation of the proposed change, the Lead Centre, in consultation with the central processing facility for that product, will develop a formal change plan for implementation across the network. This might include, for example, staggered changes across sites, the use of travelling standards to ensure consistency of changes at different sites, and preliminary analysis of the effects of the change at test sites before implementation across all sites. The formal change plan is then communicated to all sites within the network. Any changes or deviations from the documented approvals must be considered a new change and must be reassessed by the Lead Centre.

After network wide implementation of a change has been completed, the Lead Centre, together with the central processing facility for that product, will formally audit the implementation of the plan and write a report, a Change Impact Report, which be archived as part of the metadata record for that data product. The report would include an assessment of the degree to which the formal change plan was implemented. A second review should be conducted some time later (e.g. 1 year) to check for unexpected discontinuities in the measurement series.

2.3.12. Procedure for site specific change implementation

The process for justifying, accepting and implementing changes in:

i) instrument operators,
ii) location of instruments,
iii) operating environments for instruments,

will be left to the sites making those changes. Documentation of these changes in the form of metadata is essential and sites will be audited on the completeness of their metadata submitted to GRUAN archives as part of the site assessment and certification process (see Section 5.5). Sites must also provide this information to the central data processing facility for the relevant product so that these can be flagged in the metadata, which provides essential input to the data processing, as potential breakpoints in the measurement series.

3 MEASUREMENT UNCERTAINTY

3.1 Evaluating measurement uncertainty

Measurements of the atmospheric state will always differ from the true value and evaluating this measurement uncertainty is a central tenet in GRUAN's operations. A common GRUAN approach to measurement uncertainty evaluation and a common procedure to establish measurement uncertainties is required to homogenize uncertainty estimates across the network. It is also needed to ensure that the measurement and its uncertainty, where possible, are traceable to primary standards. This common definition should, ideally, also be adopted by instrument providers.

Achieving a correct estimate of measurement uncertainty may require as much, if not more, effort than making the measurement itself. However, such effort is necessary to achieve the goal of GRUAN to provide reference measurements from the surface to the upper stratosphere. The availability of an estimate of the measurement uncertainty for every measurement made within GRUAN will significantly improve the utility of the measurements, the robustness of the data, and will elevate the GRUAN measurements above what is currently available for many, but not all, measurement systems.

The availability of sufficiently detailed metadata is a vital component in evaluating uncertainty budgets. The more detailed the metadata, the 'deeper' the measurement uncertainty can be diagnosed. The approach that should be followed is that a reference standard, application of an operating procedure, use of a data processing algorithm, and all evaluated uncertainty sources must be available through the metadata tagged to that measurement. Such sources of metadata may include (Immler et al., 2010): previous measurement data, experience with or general knowledge of the behaviour and properties of relevant materials and instruments, manufacturer's specifications, data provided in calibration and other certificates, and uncertainties assigned to reference data taken from handbooks. It is vital that sources of measurement uncertainty are made transparently available to end-users of GRUAN measurements.

A particular challenge for GRUAN in evaluating measurement uncertainty is that for *in situ* measurements of upper-air ECVs, the instrumentation operates in conditions that are difficult to replicate in a controlled environment (e.g. a test chamber). Calibration of the instrument in its operating environment where e.g. transient influences of changes in solar radiation and/or clouds are likely to affect sensor characteristics is generally not possible. Furthermore, the staple instruments for much of GRUAN priority 1 variables, viz. balloon-borne sondes, are used for measurements of single profiles. The calibrated instruments with quantified measurement uncertainties are discarded after each profile measurement and re-calibration or re-characterization after a measurement is often not possible even if the instrument is recovered. The emphasis is then on employing standards that ensure stability, traceability, and uniformity between instruments and across the GRUAN network as a whole. Periodic calibrations are then required according to the stability of the sensors with respect to the prescribed target uncertainty.

Because one of GRUAN's primary goals is to detect long-term climate trends in the upper atmosphere, and because GRUAN data are likely to be used for other purposes such as satellite validation, acting as a reference for GUAN, or as input to global meteorological reanalyses, both reducing the random error in measurements (to emphasize reproducibility) and reducing the systematic error (to achieve the best possible accuracy) need to be priorities. Therefore the aim should be to identify and minimize both random and systematic errors, and to include the effects of both when calculating measurement uncertainties.

The GRUAN policy for dealing with measurement uncertainty shall be:

i) *Describe/Analyse* all sources of measurement uncertainty to the extent possible.

ii) *Quantify/Synthesize* the contribution of each source of uncertainty to the total measurement uncertainty.
iii) *Verify* that the evaluated net uncertainty is in agreement with the required target uncertainty.

3.1.1. Describe/analyse sources of measurement uncertainty

The first step in the process of evaluating a complete uncertainty budget associated with any measurement is to fully explore and describe each source of uncertainty in the form of statistical (Type-A) uncertainties and all other sources of uncertainty (Type-B). Type-B uncertainties result from the instruments used to make the measurement, the measurement procedure, the definition of the quantity under measurement, the quantities that influence the measurement, the traceability chain etc., the sensor calibration, integration, performance, and external influences to operational routines such as sensor preparation and sensor ground-checks. While a specific sensor might perform well, if its value depends on another sensor that performs less well, this source of uncertainty needs to be accounted for. For example, if an accurate temperature measurement T is made but the vertical coordinate for that measurement is derived from a pressure measurement, in the presence of large $\partial T/\partial p$, the uncertainty in pressure can introduce significant uncertainty in the temperature profile retrieval. Therefore uncertainty in the geo-location and time coordinates associated with each measurement shall also be considered when identifying and describing sources of measurement uncertainty. They become mainly Type-B uncertainties due to the knowledge (model) of the quantity under measurement, the accuracy of each of those quantities, and to the functions associating the quantities. A list of sources of measurement uncertainty will be defined in the GRUAN common approach to the measurement uncertainty evaluation. Every GRUAN site shall measure, collect, and provide all information necessary to establish an uncertainty budget for every measurement, according to the common approach.

3.1.2. Quantify/synthesize sources of uncertainty

The second step is, where possible, to quantify and correct for any measurement biases. Uncertainty in such bias corrections, which shall also be diagnosed, documented and quantified, then contributes to the total measurement uncertainty budget. Techniques to fully describe the shape of the Type-A uncertainty distributions must then be developed and higher order moments of the distribution (e.g. the skewness or kurtosis) would need to be reported as part of the measurement uncertainty description. If a measurement process can be simulated, and if the probability distribution functions (PDFs) of the various sources of uncertainty are well known, a Monte Carlo approach can be used to generate a large ensemble of 'virtual' measurements from which measurement uncertainty statistics can be calculated. This approach can be used no matter how structured or asymmetrical the individual PDFs might be. This approach has been used to estimate asymmetric errors in ozonesonde measurements (Bodeker et al., 1998). Once all biases have been corrected for, and assuming all remaining Type-A uncertainties have a known statistical distribution, the resultant net uncertainty on the measurement can be reported as a single value composed of Type-A and Type-B uncertainty components.

3.1.3. Verify measurement uncertainties

The uncertainty budget for every GRUAN measurement should be verified ideally on a sustained and regular basis but failing that at regular intervals according to any change introduced in the instrumentation or in the measurement procedure. Redundant observations from complementary instruments (see Section 6.2) can be used to verify measurement uncertainties that have been evaluated *a priori*. If coincident observations of the same ECV are available and are subjected to the same uncertainty analysis, the degree to which the measurements agree within their stated uncertainties is indicative of the validity of the measurement procedure. If measurements agree within their uncertainties, the uncertainty estimates on the individual measurements are more likely to be correct. Formal methods, based on the availability of large sets of data, have been developed to evaluate whether uncertainties are over-estimated or under-estimated (Immler et al., 2010).

For example, if two large sets of data are compared and more than 4.5% of the data are statistically significantly different within their uncertainty bars, then either a systematic effect in either or both measurement sets has been overlooked, or the uncertainties have been under-estimated, or both. On the other hand, if much less that 32% of measurement differences are smaller than the root mean square (RMS) of the uncertainties, then the measurement uncertainties have probably been over-estimated. This verification by itself does not provide a statement about the usefulness of a measurement; it only provides information about the completeness of an uncertainty analysis. Including such comparisons in operational data processing can act as a flag for where uncertainty analysis within the processing may not be complete.

GRUAN includes both *in situ* and remote sensing methods. In the case of *in situ* methods, the instrument is generally calibrated directly to the geophysical quantity of interest. This can be performed through a direct calibration generating the quantity of interest in the appropriate measurement range, or by comparison against a calibrated, usually higher class, instrument. For a number of remote sensing methods, the calibrated data are often in physical units of radiance and/or frequency, which are then analysed to provide estimates of the underlying climate variable of interest. Validation of data products, which includes verifying measurement uncertainties, is therefore a two-step process whereby the accuracy of both the instrument calibration and the analysis algorithm, are validated. Traceability of calibration is more difficult for remote sensors.

3.2 Reporting measurement uncertainty

No GRUAN measurement should be provided without an evaluation of the associated uncertainty. Where all sources of Type-A and Type-B uncertainties have been identified, and biases and deviations corrected for, including the respective additional uncertainty contributions resulting from the corrections, the measurement uncertainty can be quoted as the dispersion of the values of the measurand about the declared measured value. As discussed above, where biases remain in the measurement, or where the net random dispersion in the measurement does not follow a Gaussian distribution, alternative methods and distributions for reporting the measurement uncertainty must be considered. This may be in the form of establishing 2σ upper and lower bounds on the measurement uncertainty to denote that the uncertainty is asymmetric – generally reported as X^{+u}_{-l} where *X* is the measurement, *u* is the 2σ uncertainty in the positive direction and *l* is the 2σ uncertainty in the negative direction. Given that some systems may quote uncertainties as 1σ values, it is imperative that it is clearly stated in the GRUAN metadata that the values are 2σ uncertainties. For more complex distributions of measurement uncertainty it may be necessary to quote the most likely value i.e. the peak in the PDF for the measurement and parameters that detail the shape of the PDF (or a pointer to the PDF itself).

3.2.1. Vertical correlation of uncertainties

When comparing or combining different vertical profiles, it is important to understand the vertical correlations within each profile measurement. Each GRUAN data point is reported with its associated uncertainty. In the first instance this is done on a point-by-point basis without reference to any vertical correlation of the uncertainties. Research is required to establish the best way to determine and report these correlations.

3.3 Reducing measurement uncertainty

Changes in instrumentation or operating procedures may lead to reductions or increases in measurement uncertainty. It is important that the same, or improved, detail of uncertainty analysis is conducted for the new instrument/operating procedure as was done for the instrument/operating procedure to be replaced.

In some circumstances, e.g. in the presence of high natural variability (such as for temperature and water vapour), reducing measurement uncertainty has little impact on derived trends since the primary source of the uncertainty in the trend estimate might be the noise on the measured signal being analysed (Bodeker et al., 1998; Seidel and Free, 2006; Whiteman et al., 2011; see Section 7.4.1). It is therefore important that scientific analyses guide where reducing measurement uncertainties (both Type-A and Type-B) is most likely to lead to reductions in uncertainties in trend estimates. In the absence of instrument drift, which can be evaluated through periodic calibration, Type-B uncertainties have little effect on trend estimates, unless changes are made to the measurement process that cause systematic changes to the data.

3.4 Reducing operational uncertainty

Operational uncertainty includes uncertainties related to instrument set-up, sampling rates and the application of algorithms for data analysis. The contribution of operational uncertainty to the total measurement uncertainty in GRUAN is likely to be significantly reduced if the 'rawest' form of measurement data is submitted to a central GRUAN data processing facility (see Section 8.1) where a single, verified, validated and well described data processing algorithm is applied to the raw data. Such an approach also ensures network-wide consistency in raw data processing and resultant homogeneity of GRUAN data products. Similarly, the adoption of an identical standard operating procedure for each instrument type across the network would reduce the operational uncertainties related to instrument set-up. To this end, optimal standard operating procedures are developed at the GRUAN Lead Centre or at the site responsible for centralized processing of that ECV and then disseminated to all sites making that particular measurement and adopted where practical, with exceptions clearly documented, and agreed with the WG-GRUAN as part of the assessment and certification process.

3.5 Validating measurements

Once the uncertainty on a measurement has been calculated, the question then becomes: how well does this uncertainty represent the robustness of this measurement? Two approaches are available for validating any measurement, viz. 1) by comparing redundant measurements, and 2) by laboratory analysis of the measurement system.

When redundant measurements are present, their uncertainty must be evaluated using standardized consistency tests such as those described in Immler et al., (2010). These standardized tests must be performed across the entire GRUAN network, regardless of the type of instrument considered.

3.5.1. Comparison of redundant measurements

One way to validate a measurement is to measure the quantity of interest through two (or more) independent techniques, based on physically different measurement principles. Because the different techniques are subject to associated measurement uncertainties, comparisons yield a robust and continuous demonstration of measurement accuracy. Where simultaneous measurements of the same quantity are made using two different techniques, and disagree within their stated measurement uncertainties, it suggests that either one or both of the measurements are erroneous, or that the measurement uncertainties are under-estimated. In this way, complementary measurement techniques with different susceptibilities to local conditions can be chosen to maximize the accuracy of the data record. Additionally, uncertainty budgets validated in this way may help identify other error sources that cannot be compensated for by complementary sensors, but may be monitored *in situ*. Note that where two sensors are not measuring exactly the same volume of air due to spatial and temporal differences in the sampling and/or averaging, then an allowance for the co-location uncertainty should be included in the comparison analysis.

3.5.2. Laboratory analysis of the measurement system

The ability to generate a specific quantity or combination of quantities in the laboratory can permit an in-depth investigation of the various sources of uncertainty in the measurement. Many such facilities exist. Two examples are the environmental simulation facility at the Research Centre Juelich (Smit et al., 2007) which has provided information to validate measurement uncertainty in ozonesondes, and the radiosonde laboratory facilities available at the DWD at Lindenberg. Special chambers allowing the simultaneous generation of temperature, pressure, humidity and even wind and solar radiance can be fundamental for the specific purpose of calibrating weather instruments under the conditions of measurement. This enables the external influence factors to be determined and the matrices of variances and co-variances to be derived.

4 ESSENTIAL CLIMATE VARIABLES MEASURED IN GRUAN

The parameters most relevant to understanding changes in the climate of the upper atmosphere are temperature, pressure and water vapour. This is why, in addition to these three being the most tractable for GRUAN (see Appendix 1 of GCOS-112 and later updated in Appendix 1 of GCOS-134), they have been identified as the highest priority ECVs (GCOS-92) to be measured in GRUAN. However, to diagnose the drivers of observed changes in temperature, pressure and water vapour, a range of other ECVs also need to be measured. Therefore, a number of additional ECVs have been identified as target variables to be measured at GRUAN sites; a summary of the priority 2, 3 and 4 variables is provided in Appendix A. As scientific research into the underlying causes of observed changes in climate advances, and as the capabilities of GRUAN sites expand, this list of target ECVs for GRUAN is likely to grow.

4.1 Justification and context for essential climate variables

The purpose of this section is to provide additional scientific justification and context, and general guidelines for the measurement requirements for temperature, pressure, and water vapour. The desired measurement requirements for each of the ECVs are based on the scientific requirements of the data and not on current instrument performance, so they may not be currently achievable. In such cases the WG-GRUAN and Lead Centre will provide possible incremental approaches to achieving the target attributes for each measurement. Therefore, as stated in GCOS-134, these GRUAN requirements should be interpreted as eventual measurement goals of any site. They are deliberately aspirational being driven by likely user requirements to answer the envisaged range of scientific enquiries and were derived following exhaustive community engagement through two dedicated meetings and a solicitation through relevant list-servers.

Setting the target measurement parameters low is likely to result in stagnation since once achieved, there will be little incentive to advance. For this reason the requirements detailed below are somewhat different from those listed in the WMO/CBS requirements. The values in Appendix 1 of GCOS-134 describe what is required of the measurements to meet specific research goals and a distinction needs to be made between what is desirable and what is feasible. While they may not be currently achievable, as measurement technology advances, attaining such targets should become more likely. There are, however, many scientific objectives of GRUAN that can be achieved with current state-of-the-art capability and so in no case should a present day inability to achieve these targets result in the exclusion of a site or a measurement programme from GRUAN as long as the measurement programme is able to achieve the mandatory requirements detailed in Section 5.3. A GRUAN site shall use currently available equipment in a manner ensuring optimum performance from that system. Development and improvement of systems at GRUAN sites is to be encouraged, but these developments should be performed in a manner that does not interfere with the stability of GRUAN network observations (see Section 2.3.5).

The measurement ranges prescribed in Appendix 1 of GCOS-134 should cover the range of values likely to be encountered over the vertical range of interest so that any proposed instrument, or set of instruments, would need to be able to operate throughout that range. Measurement precision refers to the repeatability of the measurement as measured by the standard deviation of random errors (Section 2.1). However, measurement random error is closely tied to the frequency of observations since observations are often averaged and the greater the sample size, the less stringent the required precision in terms of the uncertainty on the mean. Measurement frequencies are not specified because they depend on instrument type and are also likely to vary over time. Measurement accuracy refers to the systematic error in a measurement (Section 2.1). It is not directly specified for many variables for which variations, and not absolute values, are needed to understand processes. Measurement accuracy is directly related to long-term stability, the maximum tolerable change in systematic and random errors over time, which is a critical aspect of the reference network.

4.2 Development of climate data records of ECVs

Development of climate data records of ECVs within GRUAN shall be consistent with the *Guideline for the Generation of Datasets and Products Meeting GCOS Requirements* (GCOS-143). Ensuring transparency in the generation of climate datasets and products within GRUAN is essential to enable users to judge the quality and fitness for purpose of climate datasets and products. In addition to the requirements defined elsewhere in this document for GRUAN data products, the following recommendations, consistent with those detailed in GCOS-143, are made:

i) Periodic review of climate data records produced by GRUAN should be undertaken by an external body to provide an independent assessment of its quality and thereby improve the confidence that the user community has in the product.
ii) Provide a facility for user feedback on the quality, usefulness and applicability of the data products.
iii) A quantitative maturity index describing the level of scientific maturity (1=initial, 2=experimental, 3=provisional, 4=demonstrated, 5=sustained, 6=benchmark) should be included in the description of the climate data record.
iv) A full description of the climate data record should be published in the international peer reviewed literature.

4.3 Temperature

4.3.1. Scientific justification

Upper-air temperatures are a key dataset for the detection and attribution of tropospheric and stratospheric climate change since they represent the first order connection between natural and anthropogenically driven changes in radiative forcing and changes in other climate variables at the surface. Furthermore, the vertical structure of temperature trends is important information for climate change attribution since increases in atmospheric long-lived greenhouse gas (GHG) concentrations warm the troposphere but cool the stratosphere steepening vertical temperature gradients in extra-tropical regions. Other drivers of atmospheric temperature changes, e.g. changes in solar output, would not have the same vertical profile fingerprint.

Radiosondes remain a primary workhorse within the global upper-air network for the measurement of temperature, pressure and water vapour, it is imperative that GRUAN sites establish state-of-the-art radiosonde measurement programmes that match the optimum stability of performance obtainable to date. In addition, efforts should continue to improve the quality of radiosonde measurements, where it is known there are significant limitations in performance for use in climate monitoring (WMO, 2011). Other measurement techniques can and should be developed to extend the height range of the temperature profile measurements and to reduce the random and systematic errors on the measurements. However, these should always be quantitatively inter-compared with collocated radiosonde measurements to provide a traceable link to the radiosonde measurements made within GUAN. Temperatures measured by high-quality radiosondes are needed to:

- Monitor the vertical structure of local temperature trends.
- Correlate changes in other parameters, especially water vapour (see below), with changes in temperature.
- Provide a reference against which satellite-based temperature measurements can be calibrated and adjusted so that long-term changes can be estimated globally with greater confidence.
- Validate temperature trends simulated by climate models.

- Provide input to global meteorological reanalyses such as NCEP, ECMWF, NASA, JMA and NOAA.
- Provide high quality temperature profiles for use in the analysis of remote sensing measurements of other variables.
- Provide input to numerical weather prediction models if and when submitted shortly after the measurement. Upper-air measurements of temperature and water vapour are two of the basic measurements used in the initialization of numerical weather prediction models for operational weather forecasting. In turn, feedback from the numerical analysis potentially provides a useful metadata element in the final GRUAN processed measurement (see Section 9).

The requirements for random and systematic error (bias), and long-term stability are detailed below and are guided, in part, by the needs of end-users and in particular the use of the measurements in detecting trends in temperature time series which include natural, unforced climate variability. This becomes a signal-to-noise ratio problem and climate models can be used to guide the measurement requirements given expectations of future trends in temperature and natural variability (see e.g. Figure 10.7 of IPCC 4^{th} assessment report).

It is particularly important that trends in tropical cold point tropopause temperatures are accurately detected since this is thought to control the flux of water vapour into the stratosphere (Gettelman et al., 2002; Fueglistaler and Haynes, 2005) and changes in stratospheric water vapour influence radiative forcing and temperatures both in the lower stratosphere but also in the upper troposphere (Forster et al., 2007; Solomon et al., 2010). At present, temperature trend uncertainties in the lower stratosphere and upper troposphere remain large, particularly in the tropics. For this ECV, addressing trends in tropical cold point temperatures should be a focus for GRUAN. To this end establishing close working ties between the tropical GRUAN sites at Manus and Nauru with the sites within the SHADOZ network (Thompson et al., 2007) and with the GUAN sites operating in the tropics would be particularly advantageous.

4.3.2. Discussion of specific measurement requirements

Vertical range: The effects of elevated concentrations of greenhouse gases on atmospheric temperatures are seen most clearly in the upper stratosphere (Shine et al., 2003). Vertical temperature profiles are most routinely measured using radiosondes which seldom reach above ~35 km.

Systematic error (bias): The GRUAN target for temperature bias (≤0.1 K in the troposphere and ≤0.2 K in the stratosphere) can probably be met by several of the better operational radiosondes but not in the daytime, see WMO (2011) and the revision of Chapter 11 of the CIMO Guide, published in 2012. At present,the most accurate radiosonde is possibly that using the multi-thermometer approach such as the ‘Accurate Temperature Measuring Radiosonde’ (Schmidlin, 1991) and ‘Lockheed-Martin Sippican multithermistor sonde’ (WMO, 2011) with an active radiation correction which considers different cloud and radiation environments, claiming an uncertainty of 0.3 K throughout most of the upper troposphere and the stratosphere, but this is not yet widely available in sufficient numbers for use throughout GRUAN and the documentation on the measurement method is insufficient. Thus, GRUAN should proceed with the best operational radiosondes available, using the methods of observation agreed with the GRUAN Lead Centre, ensuring that sufficient sites make a priority of temperature measurements in the dark. Development of commercially available new technology to achieve higher accuracy in the daytime is a priority.

Stability: Temperature trends over the satellite era are 0.1–0.2K/decade requiring long-term stability to be an order of magnitude smaller to avoid ambiguity.

4.3.3. Requirements consistent with state-of-the-art capability

Measurement range: 170 to 350 K, noting however that the range for which calibrations apply often does not extend as low as 170 K (see table 4.1.2 of IOM-No.107).

Vertical range: while GCOS-134 calls for 0-40 km, 0-30 km is routinely achievable with radiosondes

Vertical resolution: 100 m or better below 30 km altitude, 500 m above 30 km altitude

Random error[4]*:* ≤0.2 K

Systematic error (bias): while GCOS-134 calls for 0.1 K in the troposphere and 0.2 K in the stratosphere, 0.5 K in the troposphere and 1 K in the stratosphere are prescribed in the Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)

Stability: 0.05 K/decade

4.3.4. GRUAN measurement targets

As detailed in Section 7.2, a discussion of target measurement attributes should not occur outside of a context of a particular anticipated scientific study. In the absence of the availability of recommendations based on specific uses for the measurements, the following are provided as indicative guidelines and are taken directly from GCOS-134.

Measurement range: 170 to 350K

Vertical range: 0 to 40 km

Vertical resolution: 100 m or better below 30 km altitude, 500 m above 30 km altitude

Random error: ≤0.2 K

Systematic error (bias): ≤0.1 K in the troposphere and ≤0.2 K in the stratosphere

Stability: Better than 0.05 K/decade.

4.4 Water vapour

4.4.1. Scientific justification

Water vapour is the primary natural GHG and is central to global water and energy cycles. It acts primarily as a feedback, amplifying the effects of increases in other GHGs. Water vapour is the raw material for clouds and precipitation, and limited knowledge has compromised our ability to understand and predict the hydrological cycle, and understand its effect on radiative transfer (Peter et al., 2006). Water vapour is also a source of OH in the upper troposphere and stratosphere, influencing methane, ozone and halogenated GHGs. High clouds due to water vapour in the upper troposphere/lower stratosphere (UT/LS) affect both the planet's shortwave albedo and its longwave greenhouse effect, and both cloud particles and water molecules are involved in chemical reactions that govern stratospheric ozone concentrations. Fully quantifying the Earth’s radiation budget depends on an accurate assessment of the radiative properties of clouds and the water vapour continuum.

Changes in water vapour in the UT/LS exert a greater radiative forcing than changes elsewhere (Solomon et al., 2010). Standard radiosonde humidity sensors have generally a very poor response at the low temperatures (<-50°C), pressures, and water vapour concentrations of the UT/LS (Wang et al., 2003). Although there has been significant progress since 2003 (WMO 2011), no operational radiosonde can be expected to measure with sufficient accuracy in the lower stratosphere for climatological purposes (GCOS-112).

[4] Although ‘random uncertainty’ would be the preferable term (see Section 2.1), the term ‘random error’ is used here to remain consistent with the terminology used in GCOS-112 and GCOS-134.

A number of factors, many linked to changes in climate, are likely to affect the flux of water vapour into this climatically important region of the atmosphere, viz.:

i) Changes in the cold-point tropopause temperature (Zhou et al., 2001; Wang et al., 2012).
ii) Changes in convection. Convective transport of ice particles into the UT/LS can provide a path which bypasses the limitation imposed by the cold-point tropopause temperature (Sherwood and Dessler, 2001).
iii) Changes in the Brewer-Dobson circulation (Austin et al., 2006).

While most of the Earth's water vapour is contained in the lower atmosphere where it can be measured as absolute or relative humidity, the water vapour content of the upper atmosphere is measured in parts per million and is difficult to measure accurately; the older generation of operationally-deployed balloon-borne instruments, and the satellite data record to date did not allow the measurement of water vapour in the upper troposphere and lower stratosphere to the required accuracy to be useful for climate applications (Soden et al., 2004). However, accurate water vapour measurements in the upper atmosphere are critical, especially for radiative transfer modelling. Understanding the water vapour budget throughout the atmosphere is also necessary for interpreting measurements of outgoing longwave radiation (OLR).

Satellite-based solar occultation and limb-sounding instruments can measure water vapour in the upper troposphere and stratosphere but inter-satellite differences preclude the use of earlier data in long-term trend analyses (Rosenlof et al., 2001). Measurements of the vertical profile of water vapour with small random errors will provide valuable input data to global meteorological reanalyses and data for validating global climate models.

Instruments such as the Cryogenic Frostpoint Hygrometer (CFH; Vömel et al. 2007b), the NOAA Frost Point Hygrometer (FPH; Hurst et al., 2011b), and the Fluorescent Advanced Stratospheric Hygrometer for Balloon (FLASH-B) Lyman-alpha instrument can provide water vapour measurements in the lower stratosphere, but are very expensive compared to operational radiosondes. The Snow White chilled mirror hygrometer is able to measure reliably in the upper troposphere at night (WMO, 2011). All of these instruments require a much higher level of operator skill to ensure reliable operation compared to operational radiosondes. Where several GRUAN sites are in the same climate region, e.g. Western Europe, it does not appear justified to necessarily mandate every GRUAN site in that region to fly these systems once a month, although such operations are encouraged as they may yield new insights into instrument performance and sub-monthly and regional variability. Priority should be given to measurements in the tropics, when resources are available.

Modern operational radiosondes have much improved performance compared to those reported earlier and there has been a significant improvement between the WMO Radiosonde Intercomparison hosted in Mauritius in 2005 and that hosted in Yangjiang China in 2010 (WMO 2011). The better sensors now start to become slow to respond at temperatures around -70ºC. Some manufacturers implement a software slow-response correction which needs further investigation. A second source of error comes from assuming that the temperature of the humidity sensor in the day is the same as that reported by the radiosonde temperature sensor. However, adjustment algorithms for this slow response have been implemented and methods for reducing the solar heating effect have been implemented, so the relative humidity uncertainties in the tropical upper troposphere are very much smaller than in earlier operational radiosondes. The use of the better operational radiosondes in GRUAN will improve the capability to monitor changes in the upper troposphere, although further development of the systems should be encouraged.

Many sites are currently developing the capability to observe and analyse data from a ground-based GPS receiver, usually as part of larger international networks. These data provide continuous high-quality estimates of column water vapour which, in addition to being useful in their own right, can be used to validate vertical water vapour profile measurements from the surface up to 6-10 km; total precipitable water calculated from the radiosonde measured temperature and water vapour profiles should compare well with that measurement by the GPS receiver (Wang and Zhang, 2008).

4.4.2. Discussion of specific measurement requirements

Measurement range: the large range in values that needs to be covered by these measurements (0.1 – 90000 ppm) over a vertical range of 0 km to ~40 km presents a challenge for instrument development and operation since no single commercially available instrument is responsive over this range. Instrument packages may therefore need to include more than one instrument, each of which covers a particular region of the atmosphere.

4.4.3. Requirements consistent with state-of-the-art capability

Measurement range: 0.1 – 90000 ppm (see note above).

Vertical resolution: 50 m below 5 km and 100 m above 5 km altitude

Random error: 5% in mixing ratio in the troposphere (GCOS-134 calls for 2%) and 5% in mixing ratio in the stratosphere

Systematic error (bias): while GCOS-134 calls for 2% in the troposphere and stratosphere, WMO-No. 8 recommends 5%.

Stability: 0.3%/decade in mixing ratio and for the total column.

4.4.4. GRUAN measurement targets

The preliminary measurement targets for water vapour stated in GCOS-134 are:

Measurement range: 0.1 – 90000 ppm.

Vertical range: 0 to ~40 km.

Vertical resolution: 50 m below 5 km and 100 m above 5 km altitude.

Random error: 2% in mixing ratio in the troposphere and 5% in mixing ratio in the stratosphere.

Systematic error (bias): 2% in mixing ratio throughout the profile. 1% for total column.

Stability: 0.3%/decade in mixing ratio and for the total column.

These values are superseded by the more sophisticated analysis presented in Section 7.4 where water vapour is used as an example of the considerations required before GRUAN measurement targets for an ECV can be firmly established. The measurement target characteristics given in 7.4 are summarized in the table below.

Attribute	Trend detection		Satellite validation and radiation studies		Process studies
	Upper troposphere	Lower stratosphere	Radiance comparisons	Comparisons in retrieval space	
Vertical resolution	<1 km	<1 km	no data	< 2km	10-100 m
Systematic error	profile: 5-10%	profile: 5-10% or better	column: 3% profile: 5% in lower and mid-troposphere, 10% in upper troposphere	column: 3% profile: 10% in 2 km thick layers	profile: 10%
Random	up to 50%[5]	<10%	many comparisons: 10-20%		<10-25%

[5] For measurements made 2-3 times per week and assuming that systematic errors have been randomized using appropriate procedures.

error			individual comparison: ≤5%		
Stability	no data	no data	no data	no data	N/A
Temporal resolution	<1 hour	no data	high as possible		1 minute

If such measurement targets can be met:

i) A fully equipped GRUAN site (see Section 5.2.1) will be capable of detecting water vapour trends in the upper troposphere and lower stratosphere, validating satellite-based measurements, and conducting relevant process studies.

ii) A minimum entry level GRUAN site (see Section 5.2.2) will be capable of detecting water vapour trends in the upper troposphere.

4.5 Pressure

4.5.1. Scientific justification

Measurements of upper-air temperature, water vapour and other climate variables must be accompanied by the altitude/pressure at which the measurement is made. Data used e.g. as input to NWP primarily use standard geopotential heights, but conversion from geometric altitude to geopotential height is straightforward; converting from geometric altitude as measured for instance by a GPS radiosonde, to geopotential height does not require knowledge of the vertical temperature structure.

In most NWP models the observations are input at levels defined by the ratio of the pressure to the surface pressure. The model must therefore convert geopotential height into pressure if the system does not provide pressure observations. Deducing pressure from the geopotential height requires knowledge of the temperature and water vapour structure in the vertical profile, and if this is not directly available from the system, the model will compute the values using its own analysis fields.

If data from a pressure sensor are used to compute geopotential height for a radiosonde, the uncertainty in calculated geopotential heights will result from uncertainties in the temperature, pressure and water vapour measurements. However, most modern radiosondes now use GPS navigation signals to measure altitude and, when set up carefully, can meet GRUAN requirements for pressure/altitude observations through much of the atmosphere. In the lower troposphere, however, where current CIMO guidelines call for 0.5 hPa or smaller systematic error, this is not achievable with current GPS technology. The uncertainty in the GPS altitude measurements has very little variation with height in the atmosphere (WMO 2011). CIMO therefore recommends that GPS radiosondes are used at all GRUAN sites. However, pressure sensors are also still required to validate the GPS geopotential height measurement and to provide a pressure value for conversion of other measurements (e.g. ozone) to different units.

Pressure measurements that exhibit long-term drift, coupled with a steep vertical gradient in some target trace gas, will alias into an apparent trend in that trace gas. It is therefore essential that pressure profile measurements maintain long-term stability.

4.5.2. Requirements consistent with state-of-the-art capability

Measurement range: 1 – 1100 hPa

Vertical range: while GCOS-134 calls for 0-40 km, 0-30 km is routinely achievable with radiosondes.

Vertical resolution: 0.1 hPa

Random error: While GCOS-134 calls for 0.01 hPa random error, the recommendation from CIMO is:

1 km altitude, 1 hPa (equivalent height error of 10m)
16 km altitude, 0.3 hPa (equivalent height error of 20m)
32 km altitude, 0.05 hPa (equivalent height error of 30m)
48 km altitude, 0.01 hPa (equivalent height error of 50m)

Systematic error (bias): While GCOS-134 calls for 0.1 hPa, the WMO-No. 8 prescription is 1 hPa to 2 hPa in the troposphere and 2% in the stratosphere.

Stability: Better than a quarter of the random error quoted above, per decade.

4.5.3. GRUAN measurement targets

As detailed in Section 7.2, a discussion of target measurement attributes should, to the extent possible, not occur outside of a context of a particular anticipated scientific study. In the absence of the availability of specific recommendations based on intended uses for the measurements, the following are provided as guidelines applicable across a range of uses and are taken directly from GCOS-134.

Measurement range: 1 – 1100 hPa

Vertical range: 0 – 40 km

Vertical resolution: 0.1 hPa

Random error: 0.01 hPa

Systematic error (bias): 0.1 hPa

Stability: Better than a quarter of the random error quoted above, per decade.

4.6 Moving beyond priority 1 variables

The emphasis to date within GRUAN has been on observations of priority 1 variables. This allows testing of the guiding principles for all reference observations before expanding the measurements at GRUAN sites to lower priority variables. A fully functioning GRUAN that serves all envisaged purposes will require measurements of all ECVs listed in this section and in Appendix A of this Guide. This section of the Guide outlines the procedures and requirements for expanding the capabilities of the GRUAN network by moving beyond the priority one variables of temperature, pressure and water vapour. These procedures and requirements recognize the heterogeneity of the network and that not all target variables are likely to be observed at all sites. To achieve consistency and homogeneity of data products both at individual GRUAN sites and across the network as a whole, it is essential that the procedures developed to bring new ECVs online within GRUAN provide an end-to-end solution that details the collection of raw data and associated metadata, the processing and quality assurance of those measurements, and the provision of the data products to the GRUAN user community.

4.6.1. Requirements

For each new variable, or set of variables (e.g. cloud properties may be treated as a single set of variables), planned to be brought online within GRUAN, the following are required:

A task team: The goal of the task team is to provide the scientific basis and oversight required to bring the new variable online in GRUAN. A key task is to write the technical manuals described in Section 4.6.2 below. Membership of this task team should include one member of the GRUAN Lead Centre, at least one member of the Ancillary Measurements Task Team, a representative of the central processing facility for that ECV (see below), at least two members of the WG-GRUAN, at least one internationally recognized instrument expert for each of the instruments likely to provide measurements of the ECV of interest, and other members of the international community with expertise in the processing, quality control and interpretation of the resultant data. In some cases, more than one position on the task team may be filled by a single person. The task team is likely to remain in effect only in the lead-up phase prior to those data products flowing to users through the GRUAN data archive. The terms of reference for the task team, as with other GRUAN task teams, are defined by the WG-GRUAN in consultation with the Lead Centre. The task team is directly answerable to the co-chairs of the WG-GRUAN. Task teams are usually structured to combine instruments experts with the users of the data produced by those instruments.

A central processing facility within GRUAN: As discussed elsewhere in this guide, processing raw data collected at GRUAN sites at a single centralized processing facility is essential to ensure homogeneity of measurement time series at each GRUAN site and to ensure homogeneity of the data product across the network. The centralized data processing facility will implement the data collection, quality assurance, and processing and dissemination protocols defined in the technical documents developed in consultation with the task team.

4.6.2. Technical documents

The task team is responsible for writing, or for coordinating the writing of, the technical documents described in this section. Each document forms part of the official technical document series within GRUAN and is subject to the policies for reviews of technical documents by the WG-GRUAN.

For each instrument providing measurements of the ECV of interest, the following technical documents are required. Whether these should be standalone technical documents or combined into a single document for each instrument has not yet been resolved.

Standard operating procedures: The development of standard operating procedures for each instrument used in GRUAN is also key to achieving homogeneity of the GRUAN data product. These standard operating procedures are archived at the Lead Centre and are provided to each GRUAN site operating that instrument. Standard operating procedures for many instruments are likely to be available from partner networks (see Section 1.5) and if available should be adapted to meet the needs of GRUAN. As described in Section 5.1, while implementing these operating procedures is not mandatory, sites are required to document where they have deviated from the prescribed standard operating procedures and, when audited, are assessed for their ability and willingness to adhere to the standard operating procedures within GRUAN. The standard operating procedure technical document for an instrument includes a section describing how the instrument meets the instrument requirements in terms of information content, instrument heritage, sustainability, calibration, robustness of uncertainty, manufacturer support and site location, as detailed in Section 6.1. The document describes measurement scheduling under the guidelines provided in Section 7.2. Standard operating procedures include a detailed description of how any changes in instrument type, operating procedures, data processing algorithms, instrument operators, location of instruments, and operating environments for instruments, are managed (see Section 2.3).

Data and metadata capture: This technical document describes the process for capturing the raw data from each measurement, the metadata associated with each measurement, and the metadata associated with the measurement programme as a whole which is not measurement event specific. The requirements for the capture of raw data and metadata for each measurement, as described in this technical document, guide the development of the software tools that are developed by the central processing facility (e.g. the *RSLaunchClient* and *LidarRunClient* utilities for radiosonde and lidar data capture respectively). These requirements must be specified in complete detail including field types (scalar/vector), descriptors, units etc. It is also essential that metadata associated with the site and measurement programme as a whole, and in particular change events (see Section 2.3) that may cause discontinuities in the measurement time series, are captured. The requirements for such metadata capture, as detailed in this technical document, guide the development of the necessary tools (e.g. the *IGLIMP* for lidar metadata capture) by the central data processing facility.

Guidelines for assessment and certification: As detailed in Section 5.1, individual measurement programmes are assessed and certified for inclusion in GRUAN. This technical document defines the criteria against which that assessment and verification takes place.

Central data processing: This technical document defines how Primary Raw Data (PRD) and Converted Raw Data (CRD) streams (see Section 8.1), and including metadata, from individual sites are processed to generate Standard GRUAN Product Data (SGPD) (e.g. the *GLASS* for lidar operation in GRUAN). It includes a description of all data processing algorithms, calibration procedures and the mechanisms for ensuring traceability of the measurements to primary standards, data correction and homogenization algorithms, procedures for describing and/or analysing all source of measurement uncertainty, procedures for quantifying and/or synthesizing all sources of measurement uncertainty, and procedures for verifying measurement uncertainty (see Section 3.1). It also includes a description of the triggers that signal the need for reprocessing of historical data, either for specific sites or across the network as a whole, and how the metadata related to the measurement programme are used in this capacity. It is the responsibility of the GRUAN central data processing facility to monitor these triggers for reprocessing. A thorough description of the methods for quality control and quality assurance is also included in this technical document.

In addition to those instrument specific technical documents, the following is required:

Creation of the GRUAN data product: This technical document details any additional processing required to create GRUAN data products from SGPD. It also details how Integrated GRUAN Product Data (IGPD) are generated from SGPD. In particular, the use of SGPD to generate SASBE-type IGPD is described in this technical document. The document includes a full description of the contents and structure of the data files used to disseminate the data to users.

4.6.3. Procedures, role and responsibilities

The process of activating the generation of a new GRUAN data product begins with the WG-GRUAN, in consultation with the Lead Centre, constituting the task team, and selecting co-chairs for that group.

The Lead Centre selects the centralized GRUAN data processing facility for the ECV of interest. Sites within GRUAN are given the opportunity to volunteer for this role. The Lead Centre may also approach the site most suitable to act as the processing facility and request their participation in GRUAN in this role. The centralized facility need not be the Lead Centre or a site.

The task team develops the set of technical documents required to manage the generation of data products for this ECV in GRUAN. The development of these technical documents is done in close consultation with the central data processing facility, and the documents are reviewed under the GRUAN protocols for technical document review.

Once the technical documents have been finalized they are circulated to those sites within GRUAN proposing to provide measurements of that ECV. Of particular importance are the documents detailing the standard operating procedures for the instruments providing measurements of the ECV of interest.

Raw data and their associated metadata then start flowing from GRUAN sites to the central data processing facility followed by the generation of the SGPD and IGPD products. The task team reviews the data products before they are disseminated to GRUAN data users through the GRUAN data archives (see Section 8.6).

The central processing facility also processes historical data that might be available from GRUAN sites contributing raw measurements so that the time series of GRUAN data products are extended backward in time.

5 GRUAN SITES, ASSESSMENT AND CERTIFICATION

5.1 Introduction

The purpose of this chapter is to define the process by which sites are assessed for GRUAN certification and the process by which that certification is maintained. Certification is essential to ensure that the sites within GRUAN operate at a level that maintains GRUAN's status as a premier upper-air climate monitoring network (Seidel et al., 2009). GRUAN is more than a collection of measurements made at individual sites. Part of the scientific benefit that will accrue from GRUAN results from the homogeneity of the reference quality standard of the measurements made at network sites. A shortfall in maintaining that quality standard at one site reduces the users' confidence in measurements made across the network as a whole. Sites therefore need to be sufficiently consistent and scientifically sound in their operation for the envisioned scientific benefits to accrue. The site certification process assures that all sites operate to the same reference quality standards to guarantee homogeneity of quality across the network. This chapter provides pragmatic criteria for assessing and certifying existing sites and new site offers. These criteria are designed to be as transparent as possible and to minimize the overhead involved for all parties in the certification process.

Specifics regarding site assessment and certification include:

1. Site assessment and certification is the joint responsibility of the WG-GRUAN and the GRUAN Lead Centre. If a GRUAN site is operated at the Lead Centre it will be subjected to the same assessment and certification process as all other sites in the network. Assessment and certification of sites within GRUAN is consistent with the guidance developed with the WMO Commission for Instruments and Methods of Observation (CIMO; WMO-No. 8), the WMO Guide to Climatological Practices (WMO-No. 100) and the WMO Commission for Basic Systems (CBS; WMO-No. 488).
2. Sites seeking to become GRUAN sites will first be assessed according to their ability to meet the mandatory operating protocols defined in Section 5.3 and then according to the added value they bring to the network, as defined in Section 5.4.
3. Sites will propose specific measurement programmes for inclusion in GRUAN and it is these that will be required to conform to the operating protocols defined in Section 5.3 and which will form the basis for assessing the added value that the site brings to the network as a whole. This will enable sites to operate some, but not necessarily all, of their measurement programmes to GRUAN standards.
4. Determining whether the operating procedures for proposed measurement programmes meet the prescribed operating protocols will be done objectively against the standards outlined in Section 5.3.
5. In assessing the value which a specific site adds to the network, the WG-GRUAN will base decisions on sound scientific research while exercising its discretion in evaluating the proposal against the criteria defined in Section 5.4.
6. The Lead Centre and WG-GRUAN will provide written feedback to each site as part of the certification process.
7. To identify potential problems early, sites will be reviewed annually based on their annual reports (see below) which must highlight any operational anomalies, and based on reports on data flow, site performance etc. from the Lead Centre. More complete site audits will be undertaken every 3 to 4 years (see Section 5.6 below for more details).

If site reassessments identify measurement programmes that consistently fall short of GRUAN operating standards, GRUAN certification of that programme will be suspended. If all measurement programmes at a site lose their GRUAN certification and if jointly developed recovery plans for the measurement programmes at the site have repeatedly failed to resolve outstanding problems, the site will be suspended from the GRUAN network. The WG-GRUAN and Lead Centre will work proactively with sites to remedy these problems wherever possible in a timely and cost-effective manner.

5.2 Levels of GRUAN operation

GRUAN is a heterogeneous network that includes sites from both the research community and the operational meteorological community. Sites will vary in levels of maturity and possess varying levels of infrastructure and financial support. Some GRUAN sites will only be able to provide data to address some of the measurement objectives discussed in this Guide while others may be able to meet most or all of them. Listed below are the measurement capabilities and frequencies that might be achieved by GRUAN sites at each end of this capability spectrum. Also given are the science objectives that such measurement frequencies would address. Following from the discussion in Section 7, the scheduling details for these measurements will be based upon the actual scientific goals for each site operator and relevant local site information such as local atmospheric variability statistics, timing of satellite overpasses, balloon drift information, etc.

While all GRUAN sites will provide routine observations, some will be able to provide data in near-real-time (NRT; within 2 hours), some will be able to conduct research and development into new measurement techniques, and some will be able to do all of these. Some sites will be able to commit to a sustained multi-decade programme of measurements while other sites will have a greater emphasis on research measurements. All GRUAN sites are required to meet the mandatory requirements outlined in this section of the guide.

5.2.1. A fully equipped GRUAN site

Full achievement of GRUAN objectives will be achieved by fully equipped GRUAN sites which shall:

1) Make at least double, and preferably triple, redundant measurements of all GRUAN priority 1 and 2 ECVs[6] and, specifically:
    a. Four times daily radiosonde measurements of temperature, pressure and humidity, submitted in NRT to the WIS (WMO Information System) sufficient to achieve NWP-based QA/QC. Temperature profiles to ~30 km and water vapour throughout the troposphere. Flights either at 00, 06, 12 and 18 UTC or at 00, 06, 12 and 18 LST (local solar time), with a preference for LST[7]. In the first instance, on days when overpasses of relevant satellites will occur, the launch times of the nominal 06 and 18 UTC flights should be shifted to maximize coincidence with satellite overpass. With lower priority, where redundant measurements of temperature, pressure or humidity are available at the site, e.g. a lidar temperature profile measurement, the launch times of the nominal 06 and 18 UTC flights should be shifted to maximize coincidence with the redundant measurements. High quality surface measurements of these same variables are also required to provide a traceable link between the measurements at the lowest level of each profile. Where feasible, occasional soundings at both 00/12 LST and UTC could be used to establish clima-

[6] Vertical profiles of temperature, pressure, water vapour, wind speed and direction, and ozone. Other target variables of lower priority comprise vertical profiles of aerosol attributes including optical depth, total mass concentration, chemical mass concentration, scattering, and absorption; methane columns; surface net radiation, shortwave downward radiation, shortwave upward radiation, longwave downward radiation, longwave upward radiation, and radiances; cloud properties including cloud amount/frequency, base height, layer heights and thicknesses.

[7] 00/12 UTC observations are no longer as important for NWP since 4D data assimilation is now more common. Where higher priority considerations require sites to measure at 00/12 UTC rather than 00/12 LST, this will not count against the site.

tologies of differences, including uncertainties, which could thereafter be used to relate measurements made at one standard time to measurements made at another.

   b. At least monthly observations of the vertical water vapour profile to ~30 km. Given that high frequency natural variability in the lower stratosphere is relatively small, these profile measurements should be made when most practical and when the altitude coverage can be maximized.
   c. Hourly observations of integrated precipitable water vapour.
   d. Weekly ozone profile measurements.

2) Periods of high temporal and spatial resolution measurements capable of revealing variation of key atmospheric variables.
3) Periodic intercomparisons including other instruments used across the network.
4) A demonstrated commitment to a measurement programme extending 30 years or more into the future.
5) Fulfil all mandatory operating protocols defined in Section 5.3.
6) Fully equipped GRUAN sites are strongly encouraged, but not required, to measure priority 3 and 4 ECVs.
7) Adhere to all operational protocols defined in the series of GRUAN technical documents.

5.2.2. Minimum entry requirements

As defined in GCOS-121, radiosonde observations at GRUAN sites should consist of:

1) 1 weekly production radiosonde with the best technology currently available at the site;
2) 1 monthly radiosonde capable of capturing the moisture signal in the UT/LS and all other priority 1 variables to the best level possible with current technology, launched together with weekly radiosonde;

and should aspire to making:

3) Regular 00 and 12 LST (as a preference over UTC) launches of a production radiosonde with the best technology currently available;
4) Dual launches of sondes with highest quality humidity sensing capability in the UT/LS (flying the monthly radiosonde together with a second sonde also capable of measuring water vapour in the UT/LS)[8]; and
5) Periodic intercomparisons of a large range of sonde types.
6) At least twice daily observations of integrated precipitable water vapour.

In addition, all GRUAN sites are expected to fulfil all mandatory operating protocols defined in Section 5.3 and adhere to all operational protocols defined in the series of GRUAN technical documents.

5.3 Mandatory operating protocols

The mandatory requirements for sites reflect GRUAN's primary goal of providing reference quality observations of the atmospheric column. Reference quality observations, as defined by Immler et al. (2010), are characterised by (see also Section 2.2):

1. Calibration traceable to an SI unit or to an internationally accepted standard.
2. A comprehensive uncertainty analysis that includes all known sources of random error, has corrected for known systematic errors, and has documented those sources of uncertainty which could not be quantitatively accounted for.
3. Readily accessible documentation of the measurement process and the derivation of the measurement uncertainty with a preference for publications in the peer-reviewed literature.

[8] Added by WG-GRUAN after formal workshop close

4. Validation of the measurement and its uncertainty e.g. through intercomparisons with redundant observations.
5. Availability of complete metadata which provides sufficient information to fully describe the context of the measurement. This necessarily includes the raw data and sufficient details of the processing chain.

The emphasis is on *how* the measurements are made rather than specifically on *what* measurements are made. These requirements define GRUAN's unique nature while accommodating the diverse capabilities of sites within the network. These protocols also recognize that GRUAN is not the sole stakeholder at any of the sites. Therefore, sites shall:

1. Provide reference quality observations as defined above. In particular every measurement must be traceable to fundamental standards and calibrations through well documented routines.
2. Provide uncertainty estimates for each datum or collaborate with other sites, instrument developers, GRUAN Task Teams and the GRUAN Lead Centre to provide these estimates in a consistent manner for a given instrument across the network. Profile measurements require uncertainty estimates for each measurement point on the profile. Documentation describing the calibration methods applied to each measurement, and the sources of measurement uncertainty excluded and included in the uncertainty estimate, must be provided.
3. Provide access to raw data and assure long-term storage of the raw data either at the site, at another GRUAN facility, or at another internationally accessible archive in accordance with the GRUAN Data Policy document.
4. Provide complete metadata for each measurement as defined in the requirements documents developed by the Lead Centre[9]. Metadata need to be sufficient to allow reprocessing of raw data by an independent party and will depend on the measurement system employed.
5. Provide traceable ground/instrument checks at the time of each profile measurement, independent of the manufacturer, for any instruments which provide vertical profiles extending from the surface.
6. Provide calibration information about the measurement systems (*in situ* and remote sensing) on timescales sufficient to diagnose changes in measurement uncertainty arising from changes in measurement system and in the calibration procedure.
7. Provide redundant reference observations of the ECVs selected for measurement at the site at intervals sufficient to validate the derivation of the uncertainty on the primary measurement (noting that this validation is generally achieved through comparison against other recognized reference observations).
8. Provide annual reports summarizing GRUAN operations at the site, the extent to which standard operating procedures developed for the network as a whole have been adhered to, changes in instrumentation, how those changes were managed, improvements made, progress towards achieving NRT data submission etc. Present these reports at the annual GRUAN meeting.
9. Conduct measurement programmes with an operational philosophy of continually striving to improve measurement accuracy. Actively conduct research through intercomparisons, laboratory studies, work with other GRUAN sites and/or cooperation with manufacturers to improve measurement accuracy.
10. Manage change proactively as defined in Section 2.3.
11. Participate actively in the work of the task team of site representatives. Have a site representative on this task team and a reserve contact for GRUAN purposes

[9] http://www.wmo.int/pages/prog/gcos/index.php?name=Manualsinstruments

12. Actively communicate with the Lead Centre, WG-GRUAN, Task Teams and/or other sites, (e.g. through attendance at meetings, blog postings etc.).

These mandatory operating protocols do not replace the target measurement requirements (accuracy, stability, etc.) defined in GCOS-112 and GCOS-121, which remain the targets for GRUAN. The mandatory operating protocols detailed here rather emphasize the importance of how the measurements are made, and in particular what is required to guarantee reference quality observations, rather than what physical measurements are made.

5.4 Criteria for assessing added value

Once a site has committed to operating a set of measurement programmes under the protocols defined in Section 5.3, the added value that a site brings to the GRUAN network will be assessed according to:

1. The extent to which a site can fulfil the measurement programmes expected of a fully equipped GRUAN site (Section 5.2.1). Achieving each of these measurement programmes is not mandatory for the inclusion of a site in GRUAN. However, the extent to which a site can meet these requirements will determine, in part, the additional value that that site brings to the network. While weekly sampling significantly underestimates monthly standard deviations in temperature, differences between detectable trends for weekly sampling compared to twice-daily sampling may be acceptably small (Seidel and Free, 2006). So, for example, a site that makes weekly reference quality radiosonde measurements of temperature, pressure and humidity in a large region of the globe containing no other GRUAN sites might be assessed as adding as much value to the network as a site making twice-daily reference quality measurements but located very close to another site making the same measurements. These high priority measurement programmes will be refined as the research which forms their basis progresses. This documentation will be updated to reflect these scientific advances.
2. The extent to which the site measurement programmes provide measurements in regions, or of atmospheric phenomena which were not previously sampled. In this case, the added value will depend on the locations and capabilities of the sites already participating in the network.
3. The extent to which a site brings unique observational and/or analysis capabilities aligned with GRUAN scientific objectives to the network as a whole and the likelihood of being able to propagate those capabilities across other sites in the network.
4. The extent to which a site is prepared to forgo locally established operating procedures and adhere to the standard operating procedures established by the Lead Centre and adopted by the majority of the sites already in the network. Unwillingness or inability to do this would count against a site in the assessment of the added value it would bring to the network.
5. The availability of historical measurements that conform to the GRUAN standard. All else being equal, a site that extends an existing multi-decadal time series of reference quality measurements will be assessed as adding more value to the network than a site that would initiate the same measurement programme starting at the present. Detailed documentation would be required describing how changes in standard operating procedures, instruments, calibration procedures, data processing algorithms and operators over the history of the measurement programmes have been managed to ensure that the historical measurements are reference quality. Where historical reference quality measurements are available, consideration will be given by the WG-GRUAN and Lead Centre to providing these as GRUAN data through the GRUAN data archives.
6. The extent to which a site can commit to a multi-decade programme of measurements. While it is recognised that a multi-decade programme of measurements cannot be guaran-

teed, a statement of intent with documented support (e.g. from the host institution or relevant funding agency or the PR of the country) will add to the assessment of the value that the site brings to the network.

7. The extent to which a site can provide redundant observations of the priority 1 variables (temperature, pressure, water vapour) or can conduct periodic intercomparisons of a large range of instrument types.
8. The extent to which a site is capable of measuring other ECVs identified in GCOS-112 as being desired quantities.
9. The level of institutional support for the site and commitment to maintaining long-term reference quality measurement programmes. If, in addition, a site can demonstrate that it is actively pursuing resources to enhance its capability, such as the addition of new measurement programmes, this would also enhance the added value the site would bring to the network. It is also desirable that there is full host institution commitment to GRUAN-related activities and that this commitment is not dependent on a single individual.
10. The level of institutional support for the site (and any partner institutions) to undertake fundamental scientific research of the measurements from the site and other GRUAN sites. Because GRUAN includes aspects of both operational and research networks, a strong and ongoing science programme is required to ensure that GRUAN fulfils its role as a research network.
11. The degree of historical or planned cooperation with other sites both within and outside the GRUAN network including other GRUAN-relevant networks e.g. NDACC, BSRN, GAW and GUAN.
12. GRUAN will require a minimum number of sites that can maintain a sustained measurement programme that meets GRUAN’s goals. Sites that can commit to a programme of sustained measurements will be assessed as having higher value than sites that cannot.

Such assessments of added value rely on the expert judgement of the WG-GRUAN and Lead Centre, recognize the heterogeneity of the sites within the network, and facilitate a practical approach to expansion of the network following the 2009-2013 implementation phase for GRUAN (GCOS-134) and it amendments.

5.5 The assessment and certification process

A schematic of the site assessment and certification process is provided in Figure 2. Proposals for the addition of new sites to GRUAN are likely to happen through two possible routes, viz.:

- The WG-GRUAN and/or Lead Centre invites a site to become a GRUAN certified site.
- An external organization (e.g. a national meteorological or hydrological service) approaches the Lead Centre or WG-GRUAN to propose a site.

Once a site has been identified for possible inclusion in GRUAN, through either of the routes listed above, the following sequence of events will be used to assess the site for potential GRUAN certification:

1. Provision of the GRUAN manual and this document, guidelines for the operation of specific instruments in widespread use in GRUAN, as well as documentation describing data submission protocols and the procedures that must be followed when data are submitted to the internal GRUAN archives, to the candidate site by the Lead Centre.
2. The response from the candidate site should include:
   a. A list of the measurement and calibration programmes at the site proposed for inclusion in GRUAN. This need not necessarily include all measurement programmes at the site. If a new or existing measurement programme is later proposed for inclusion in GRUAN, a similar procedure to that defined here will be used to include that programme in the GRUAN certification for that site.
   b. A complete description of how those measurement programmes will be conducted. Such information would include, for example, detailed standard operating proce-

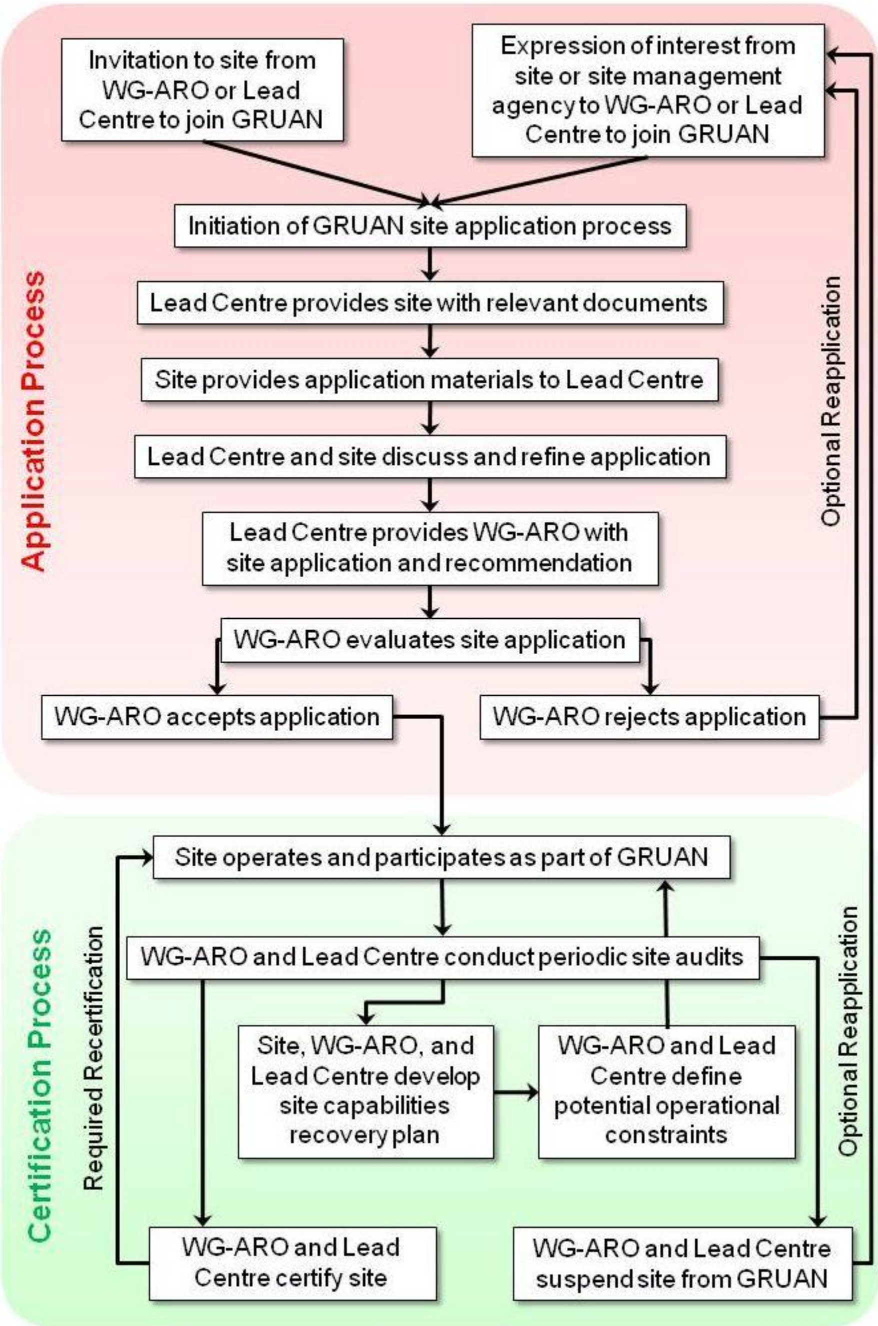


Figure 2: Schematic representation of the site assessment and certification process.

dures for each of the measurement programmes, including a description of data storage policies, and a description of how random and systematic errors in the measurements will be derived and reported together with a detailed traceability scheme. This information must be sufficient to establish the ability of the site to meet the mandatory operating protocols detailed in Section 5.3.

   c. For measurement programmes for which a GRUAN data product has not yet been well defined, the site must describe their intended strategy for developing the existing observational product into a GRUAN data product that fulfils the mandatory operating protocols defined in Section 5.3. In such instances, cooperation with other sites already in the network is highly desirable to ensure that this expertise is disseminated to similar measurement programmes in operation at other sites.
   d. The management structure of the site and a general description of the manner in which the site is operated. This would include a description of current and expected future funding levels for on-going operation of the site.
   e. A description of which data centres the measurements have previously been submitted to and are currently being submitted to.
   f. A description of how past measurements from the site have been processed. This will be used to assess whether the time series to date meet the standards for a GRUAN reference measurement. Particularly important in this regard will be detailed documentation around how changes in standard operating procedures over the history of the measurement programmes have been managed to derive a homogeneous time series of measurements. Since the historical database of measurements is an important aspect for assessing the added value that a site brings to the network (see Section 5.4), it is particularly important that the historical data can meet the stated GRUAN requirements for long-term homogeneity.
   g. A list of the scientific experts employed at the site who would likely participate in the analyses of the data collected within GRUAN. This may include mention of experts at partnering scientific organizations.
   h. Any additional information required to assess the site against the requirements listed in Sections 5.3 and 5.4.
3. There is likely to be some iteration between the Lead Centre and the candidate site to confirm specific details, fill in information gaps, and finalize the documentation from the candidate site.
4. Based on the documentation received from the candidate site, the Lead Centre will then write a short recommendation. This, together with the documentation from the candidate site, will then be submitted to the WG-GRUAN who will evaluate the proposal within 6 calendar months against the requirements listed in Sections 5.3 and 5.4. One or more visits to the site by members of the WG and/or Lead Centre within this 6 month period may be required to obtain specific additional information about the measurement programmes slated for inclusion in GRUAN at that site. If accepted, these measurement programmes will then be included in the GRUAN certification for the site.
5. Regardless of the outcome, the WG-GRUAN and Lead Centre will provide written constructive feedback to the candidate site outlining strengths and weaknesses of their programme for GRUAN purposes and suggestions as to future improvements for GRUAN operational purposes. This feedback is non-binding but rather intended to provide useful guidance and support to site capability development and retention of current capabilities.

Sites currently within GRUAN, including the site at the Lead Centre, will be assessed and certified in a similar manner.

5.6 Site auditing

Certification of GRUAN sites will not be a single event. Periodic (every 3-4 years) complete auditing of the measurement programmes included in the GRUAN certification for a site will be conducted to ensure that the programmes continue to meet GRUAN standards. Such an audit may include:

1. A review of annual reports from sites on GRUAN activities.
2. A written report from the site – essentially an update of the original report written to initiate the assessment and certification process.
3. A site visit by selected members of the WG-GRUAN and the GRUAN Lead Centre. Such a visit would include discussions with the scientists responsible for the measurement programmes at the site.

It is important for external perceptions of GRUAN integrity that these audits are conducted by the WG-GRUAN and Lead Centre and not based exclusively on annual site reports. In the eventuality of identified site problems the following protocols will be followed:

1. Should a measurement programme at an existing GRUAN site show significantly reduced observational capability over more than a year, as evaluated by the criteria listed above, the WG-GRUAN and Lead Centre will investigate the circumstances at that site, and, if needed, exclude that programme from the GRUAN certification for that site. The WG-GRUAN and Lead Centre will work proactively with sites to resurrect such programmes providing technical and in-kind support as practical.
2. Should the overall contribution of a site be deemed sufficiently diminished to call into question its continued presence in the network, the site will be informed immediately in writing. The site will be given six months to form a capabilities recovery plan, in consultation with the Lead Centre and WG-GRUAN. Should this plan be accepted the site will have no more than two calendar years from its acceptance to implement agreed key aspects. In the eventuality that this is not achieved, the site will be suspended with an invitation to submit anew at such a time as problems are remedied.

An existing GRUAN site may also request the temporary suspension of some or all of the measurement programmes at that site from GRUAN certification. This could occur for example in case of unforeseen budget limitations, non-availability of personnel or some other unavoidable circumstance affecting the measurement programmes at the site. Such a request must be submitted in writing to the WG-GRUAN and the Lead Centre. At some later time, should the site request recertification of those measurement programmes previously suspended, the procedure for certification as outlined in Figure 2 will be followed.

Along with the cooperation and goodwill of participating sites, nations, and individuals, the establishment of these GRUAN site assessment and certification guidelines provides one of the main foundations for ensuring that GRUAN meets its goals as a climate observing network.

6 INSTRUMENTATION

6.1 Instrument selection

Periodic review of instrumentation likely to be of use within GRUAN shall be undertaken since instrument technology is constantly evolving. It must also be recognized that not all sites within GRUAN will operate the same instrumentation, e.g. a new site may decide to adopt the most recent technology while a site that has a multi-decade record using an older instrument may decide to continue to use that instrument to avoid potentially introducing a discontinuity in the measurement time series. In any event, GRUAN will not prescribe the use of specific instruments in the network since the emphasis is not on prescribing an instrument, but rather on prescribing the capabilities required of an instrument and allowing individual sites to select an instrument that achieves those capabilities. That selection is also likely to be influenced by other scientific, programmatic, and practical constraints on the site. That said, the fewer the number of different types of instruments and measurement techniques deployed within GRUAN, the more likely network homogeneity will be achieved.

A number of factors should be considered when selecting instruments for use in GRUAN including (Immler et al., 2010):

- *Information content*: Are the temporal and spatial resolution, dynamic range, and other characteristics of the measurements made by the instrument as consistent as is technically feasible with GRUAN requirements?
- *Instrument heritage*: How long has an instrument been in use by the community and for what purpose? In what other networks is the instrument deployed? How substantial is the body of literature documenting its performance and measurement uncertainty? How widely distributed is the knowledge base that facilitates the instrument's successful operation?
- *Sustainability*: Are the costs for operating the instrument and the demands on personnel for operating the instrument consistent with the resources available at GRUAN sites? Is the commercial demand sufficient, and the technology available, to support the production and use of the instrument for a sufficient length of time for the expected multi-decade deployment within GRUAN? Does the instrument manufacturer plan to continue production into the foreseeable future, even if a newer (but not necessarily better) instrument is developed and marketed?
- *Robustness of uncertainty*: Is the underlying accuracy claim for the instrument and its resultant data sufficiently robust i.e. is it likely to be able to meet the uncertainty and stability standards (see Section 4.1) required by GRUAN?
- *Manufacturer support*: Is the manufacturer committed to a process of improving the performance of the instrument? Is the manufacturer prepared to actively participate in instrument intercomparisons? Is the manufacturer willing to disclose the necessary information required to form a fully traceable chain of sources of measurement uncertainty given that in some cases this information may have to be kept from public display by GRUAN beyond trusted processing individuals with a restriction to bona fide research purposes so as not to undermine the competitive advantage of the manufacturer? For a consistent uncertainty analysis it is imperative that the algorithms used for corrections within the data processing software are made available by the instrument manufacturers to those conducting the uncertainty analysis. This may be a small group of people who have signed a non-disclosure agreement with the manufacturer to protect the manufacturer's intellectual property. The fundamental requirement is that the information required to reprocess the data at any time in the future must be made available (though not necessarily publically available).

- *Site location*: Instrumentation may have to differ by climate region. For example, high-latitude sites exhibit extremely low water vapour contents in winter compared to equatorial sites. Therefore, instruments such as water vapour radiometers operating at 23.8 and 31.4 GHz, which have limited sensitivity for integrated water vapour amounts below 5 mm, would need to be augmented with more sensitive microwave radiometers operating near 183 GHz.

6.2 Measurement redundancy

Having different instruments at GRUAN sites measuring the same atmospheric parameters will be invaluable for identifying, understanding and reducing systematic effects in measurements. One of the goals of GATNDOR is to quantify the value of redundant measurements and assess optimal combinations of measurements. If successive reductions in measurement uncertainty with the addition of each coincident measurement from a different instrument can be quantified in a scientifically robust way, this provides a powerful justification for measurement redundancy at GRUAN sites. It should also be noted however that not all instances of measurement redundancy are equal. Some combinations of instruments may be more useful than others both in terms of reducing measurement uncertainty but also for generating a more complete or valuable representation of the vertically resolved time evolution of the ECV of interest.

Studies to quantify the error reduction resulting from increasing redundancy of measurements require an assessment of the uncertainty of the measurements using each of the considered techniques and then an investigation of possible sensors' synergies to reduce the uncertainty. Such studies should focus on the most common instruments in use at GRUAN sites. The quantification of the value added by complementary observations should be assessed with respect to:

- Sensor calibration/inter-calibration (here the ARM Value Added Products could be considered as a model).
- Identification of possible biases.
- Representativeness of measurements i.e. which horizontal and vertical region of the atmosphere the measurement represents.
- Quality control/assurance with a focus on instrument performance in different meteorological conditions.

The final goal of these studies is to provide recommendations for an optimal observing strategy, reducing systematic errors of measured parameters and reducing uncertainties through redundancy. Such studies are required to highlight the value of having multiple measurements of the same climate variables at GRUAN sites.

One important factor for GRUAN is that redundant measurements of the same (or related) variables should be reported in a consistent way. The cross-checking of redundant measurements for consistency should be an essential part of the GRUAN quality assurance procedures. Since all data are to be reported with uncertainties, a consistency check is, in principle, a straight forward task.

6.3 Surface measurements

While GRUAN is, by definition, an upper-air network, surface measurements at sites should also be made in such a way that:

- They are made according to WMO guidelines (WMO-no. 8), including traceability to SI standards. The CIMO classification for stations should be applied.
- The surface measurements provide ground-truthing for vertical profile measurements. For example, comparisons between ozonesonde measurements of ozone at the surface against a

high precision standard, provides essential information for quantifying uncertainties in the ozonesonde measurement.

- The measurements can, where relevant, constrain retrievals applied to remotely sensed profile data. Some remote sensing instruments that derive vertical profile data from e.g. optimal estimation techniques (Rodgers, 2000), are better constrained when a surface measurement with small random uncertainty is included as input to the forward model used in the retrieval. In some cases remote sensing of column amounts of a trace gas can benefit from having collocated surface measurements of that trace gas e.g. as is done in TCCON (Total Carbon Column Observing Network).

Until formal requirements for GRUAN sites to include surface measurements are developed, the interim guidance is that where such measurements would significantly add to the quality or utility of the GRUAN measurements, these surface measurements should be made.

6.4 Upper-air measurements

6.4.1. *In situ* instruments

As discussed in Section 4.3.1, the fact that radiosondes will remain the primary workhorse within both GUAN and GRUAN for the measurement of vertical profiles of temperature, pressure and humidity and are not recoverable, has important implications for GRUAN operations, viz.:

- The temperature sensors are not usually the limiting expense in the cost of a modern operational radiosonde and good sensors can be obtained relatively cheaply. High quality humidity sensors, on the other hand, may incur additional cost. The exposure/mounting of the sensors on the radiosonde is a limiting factor on the performance of many radiosondes, so there is still scope for improvement with the current systems without investment in very expensive replacement technology.
- Maintaining long-term stability in a radiosonde measurement time series is challenging when the instrument being used to make the measurement is discarded after each measurement. Each instrument must be individually calibrated and tied to common calibration standards to ensure long-term stability, traceability and comparability. It must also be able to retain its performance throughout an ascent, and currently this is one of the limitations of the best operational radiosondes where the bias does not appear stable to 0.1 K during an ascent (WMO 2011). The better manufacturers have managed to eliminate most faults that occur through production engineering, but any given radiosonde type has shown small fluctuations in performance with time, when checked on the ground, although these variations in performance during flight should be minimised by the ground check procedures used.

6.4.2. Remote sensing instruments

Most remote sensing instruments currently available for the measurement of priority 1 and 2 ECVs in the troposphere and lower stratosphere can be considered to be ‘research grade’ instruments. Remote sensing instrument types shall be selected for use in GRUAN based on either one, or both, of the following criteria:

1) They are recognized to be providing quality measurements of priority 1 or 2 ECV in the troposphere and/or lower stratosphere to the extent that these measurements may be considered reference measurements.

2) They are recognized to be providing valuable complementary contributions to the priority 1 or 2 ECV GRUAN *in situ* measurements (including measurement redundancy).

The ground-based remote sensing techniques currently identified as meeting one or both the above criteria are: lidars, microwave radiometers and spectrometers (MWR), GNSS-PW (Global Navigation Satellite System Precipitable Water), and Fourier-Transform Spectrometers (FTS). All four techniques have shown significant, complementary contributions to *in situ* measurements as they can all provide continuous (and/or integrated) measurements over extended periods of time. Balloon-borne *in situ* measurements are usually regarded as instantaneous at one given altitude and time, while the lidar, MWR and FTS instruments can provide several, uninterrupted hours of measurement at one given location. They therefore represent useful complements to balloon-borne *in situ* measurements since they can bridge sampling intervals between consecutive balloon launches. They are ideal instruments for process studies on timescales ranging from minutes to hours, i.e., timescales that cannot be resolved by individual balloon-borne instruments. Furthermore, the same instrument can be operated for periods of up to several decades (noting that parts may need to be replaced, requiring change management) to produce long-term homogeneous time series. Hence, ground-based remote sensing instruments provide useful information for the homogenization of time series measured by other techniques.

Due to the large variety of techniques and species involved, it is not possible to enumerate all possible combinations of *in situ* and remote sensing instruments that would be suitable for GRUAN. Generally speaking, the lidars provide high vertical resolution profiles (a few tens of metres) while MWR and FTS provide lower vertical resolution profiles (typically 3 to 6 km). However, lidars usually require more maintenance and operational overhead than MWR and FTS. The strengths and weaknesses of each technique are described in their respective individual *Best Measurement Practices and Guidelines* technical documents. An overview of the characteristics of these techniques is provided below.

Lidars

Rayleigh lidars currently provide night-time measurements of ozone (Differential absorption lidars or DIAL) and temperature in the stratosphere and night-time and daytime measurements of ozone (DIAL) in the troposphere. Vibrational-rotational Raman lidars and DIALs provide night-time and daytime measurements of water vapour in the troposphere and occasionally the lower stratosphere, and pure-rotational Raman lidars provide temperature measurements in the troposphere and lower stratosphere. In most cases, daytime lidar signals will contain significantly more noise than night-time lidar signals due to the background solar radiation, and their vertical range during daytime will therefore be much reduced. Lidars provide highest quality measurements under clear-sky conditions. Measurements are still possible in the presence of thin clouds, but are precluded above any moderately-thick cloud layer. With instrumental sampling of the order of a few meters, lidars can resolve very fine vertical structures as is the case for *in situ* measurements. At short integration times (i.e., a few minutes at most) they provide a purely Eulerian view of the atmosphere not available at such high vertical resolution with any other *in situ* or remote sensing techniques. Raman lidars need to be calibrated, which for all cases except the so-called 'first-principle' calibration, is performed on level 2 data. If the instrumentation is not interfered with, re-calibration may only be needed on a monthly or possibly yearly basis. If the instrumentation has been interfered with, re-calibration will likely be needed just after the modifications. Details on the lidar technique can be found in the *GRUAN Lidar Guidelines* technical document.

Microwave radiometers

Microwave radiometers are passive instruments measuring the down-welling natural emission from the Earth's atmosphere. The microwave receivers are calibrated to measure atmospheric radiance (often converted to brightness temperature) from which estimates of some atmospheric thermodynamic properties are retrieved.

The atmospheric parameters that can be retrieved depend on the channel specifications of the operating unit. Channels in the 22-35 GHz band provide measurements for retrieving water vapour and cloud liquid water. Two channels (usually 23.8 and 30-31 GHz) are sufficient to retrieve the column integrated water vapour (IWV) and integrated liquid water (ILW) simultaneously. Additional channels provide information on the vertical distribution of water vapour content, though at low resolution (~2-3 pieces of independent information in the troposphere) due to heavy information redundancy.

Channels in the 50-60 GHz band provide measurements for retrieving atmospheric temperature profiles in the troposphere. Temperature profiles can be estimated either by single-channel observations at several elevation angles or by multi-channel observations at one or more elevation angles. Most of the information on the vertical temperature structure is in the lowest 1-2 km. Elevation scanning is useful for increasing the vertical resolution of temperature profiles in the planetary boundary layer.

The most common retrieval types are: statistical regression, where brightness temperatures are correlated with the parameter under study (IWV, ILW, water vapour or temperature profile), a neural network based on a set of radiosonde measurements and corresponding calculated brightness temperatures, and optimal estimation where a cost function is minimized. In each case *a priori* knowledge is required. The retrievals are reported together with the *a priori* information as well as the averaging kernel functions which characterize the vertical resolution and the sensitivity of the retrieval.

Finally, instruments with channels in both the 22-30 and the 50-60 GHz bands are often called microwave radiometer (humidity and temperature) profilers. Instruments operating in the 20-60 GHz range can perform under all-weather conditions, though the quality of retrieved atmospheric parameters degrades under conditions of precipitation.

Microwave spectro-radiometers

Microwave spectro-radiometers are microwave radiometers equipped with a spectrometer that is capable of spectrally resolving the pressure broadened emission line of water vapour (e.g. 22.2 GHz, 183.3 GHz). Most instruments are equipped with digital FFT-spectrometers that have a total bandwidth of up to 1 GHz and spectral resolutions as good as 10 kHz. Being a passive technique, observations can be performed day and night except under conditions of precipitation. By combining information from the measurement and *a priori* information, it is possible to use an optimal estimation technique to retrieve water vapour profiles from ~25 km to the mesopause. The upper limit is given by the altitude where Doppler broadening begins to dominate pressure broadening whereas the lower limit results from instrumental artefacts and restrictions given by tropospheric humidity. Altitude resolution of this technique is on the order of 5-10 km. Averaging kernels, essential for a proper interpretation of humidity profiles by microwave spectro-radiometers, should be provided for all instruments used within GRUAN.

Fourier-Transform spectrometers

Ground-based Fourier-Transform Infrared (FTIR) spectrometers record infrared solar absorption spectra at a high spectral resolution (up to 0.002 cm^{-1}). The observational line of sight (LOS) through the atmosphere follows the path of the sun. Observations are limited to clear-sky conditions. The technique can simultaneously measure many different trace gases since it can produce broadband spectra in the mid to near infrared region where many atmospheric species are energetically active. The retrievals apply the differential absorption principle, i.e. there is no need for an absolute calibration of the instrument measuring the radiances. Furthermore, the spectra are measured at high signal-to-noise ratio (up to 2000). As a consequence, total column amounts can be detected with small random error. For water vapour, the random error is better than 1-2% (this is a result from theoretical as well as empirical uncertainty assessment studies).

For profile retrievals, an accurate knowledge of the instrumental line shape is required. The instrument line shape is estimated every few months by measuring the transmittance spectrum of a standard low pressure cell. The profile retrievals are accompanied by an averaging kernel documenting the vertical resolution and sensitivity of the system. Typically for water vapour, 3-4 independent atmospheric layers can be resolved. The vertical resolution of the profiles depends on the total water vapour column and is slightly higher for lower water vapour columns. Typically it is 3 km close to the observation site, 6 km in the middle troposphere, and 10 km in the upper troposphere. Initial theoretical and empirical error assessment studies indicate that the random uncertainty of the profiles is better than 20%. The spectroscopic line parameters (HITRAN), applied in the forward model, are the dominant source of systematic uncertainties.

Retrieval of other gases of interest (e.g. CO_2, CH_4, N_2O, O_3 and other trace species) can be performed with similar techniques. Characteristics of the retrieved data (uncertainties, vertical resolution) are dependent on each species' unique infrared spectrum.

6.5 Instrument co-location

Some of the current GRUAN sites, and many potential sites, consist of instrument clusters spread over some region rather than single compact sites. Some of them are in geographical locations that have complex orography and/or heterogeneous surface characteristics. There remain open questions about how physically far apart measurements can be made and still represent a GRUAN site measurement of a single column. Co-location requirements for different variables and instruments are begin developed by GATNDOR, based on sound scientific analyses, and will form the basis for deciding, as part of the site assessment and certification process (Section 5.5) whether instrument clusters meet the co-location requirements for GRUAN.

6.6 Calibration, validation and maintenance

6.6.1. Instrument calibration

Establishing reliable calibration procedures for the instruments being used within GRUAN, and applying these uniformly across the network, is an absolute prerequisite for achieving the GRUAN goals. In addition to establishing calibration procedures at individual sites that minimize the uncertainty introduced into the measurement chain (see Section 2.3) and avoid introducing discontinuities into the time series, it is equally important that site-to-site differences in calibration procedures do not compromise the goal of achieving homogeneity across GRUAN as a whole so that a measurement of a given parameter at one site is directly comparable to a measurement of the same variable at a different site. A guiding principle that will achieve this goal is that when two identical instruments are deployed at two different sites, they shall also use the same calibration procedures, preferably traceable to the same standards, and should also employ identical data processing algorithms. Exceptions to this will need to be clearly documented and the impact appropriately accounted for in the uncertainty analyses. While achieving a common data processing for each instrument will be facilitated through processing the raw data at a single central data processing facility (see Section 8.1), the same approach cannot be used for calibration procedures. To this end, achieving inter-site homogeneity would be improved for some measurement systems by developing travelling calibration standards, where possible, which can be taken to different GRUAN sites to be used in on-site calibration by comparison or inter-comparisons, as advised by the relevant task team. This is one option, but task teams should explore the best method for each measurement technique. A current example of this would be Dobson Spectrophotometer #83 which is used in the NDACC and WMO/GAW networks to achieve homogeneity across the global Dobson network (see Sections 1.5.4 and 1.5.6).

Traceability to recognized measurement SI standards that can be reproduced globally and over long periods of time will be the key practice that enables GRUAN to provide reference measurements useful for long-term climate observations. Traceability is a property of a measurement that is manifest by an unbroken chain of measurements to a recognized standard, with fully documented uncertainty at each step. This then allows a robust calculation of the propagation of uncertainties from the fundamental standard to the final measurement. If traceable standards are available across the GRUAN network, this will support achieving coherence across the network.

GRUAN sites shall maintain a "GRUAN site working standard" for each fundamental measurement unit, e.g. a thermometer periodically calibrated to a National Metrology Institute or other accredited agency standard since this ensures traceability to an SI standard. A mechanism shall be implemented to address the compatibility of those systems with the rest of the network that may not be traceable to SI standards. However, the goal is to ensure that all measurements within GRUAN are traceable to SI standards.

The use of calibrated instruments will also aid operators to detect and quantify systematic errors in GRUAN measurements (see Section 3.2). Where the final data product of a reference observation depends on ancillary measurements, these measurements must again be traceable to standards. Traceability will also permit the network to incorporate new scientific insights and new technological developments, while maintaining the integrity of the long-term climate record. To achieve traceability, metadata on all aspects relating to a measurement and its associated uncertainty shall be collected. Each site shall maintain accurate and complete metadata records and provide these to the GRUAN archives. Copies of calibration certificates shall be submitted to the GRUAN metadatabase.

The schedule of field recalibration and validation procedures should be drawn initially from experience with a given sensor type, then refined according to the results of laboratory tests and intercomparisons. The date and nature of field recalibrations should be included in metadata, so that if future experiments reveal shortcomings in schedules or methods that were in use, uncertainty estimates can be adjusted after the fact to reflect that newly-discovered information.

6.6.2. Instrument validation

Validation of the instruments used within GRUAN should include well documented and traceable calibration procedures, participation in regular intercomparisons with similar instruments used at other sites and/or intercomparisons with a travelling standard, and operational comparison of uncertainty estimates on the resultant measurements with those from other instruments (see Section 3.1.3). Most sites will likely not have identical instrumentation, with the result that instrument validation will often be site specific. A standard recommendation for the use of redundant instrumentation and remote sensing instrumentation should be developed by the Lead Centre (in consultation with GATNDOR and task teams) to aid site specific, regularly scheduled, instrument validation. The purpose is to make sharing and communication of best practices across sites seamless and continuous.

6.6.3. Instrument and site maintenance

GRUAN sites are equipped with sophisticated, state-of-the-art instrumentation and should comply with strict requirements of site maintenance, exposure of instruments and calibration performance to avoid degradation of the quality of the measurements. To ensure that the goal of long-term high quality climate records is reached, site scientists who are leading experts for the instruments used at the respective GRUAN sites shall take responsibility for individual instruments operated at the

GRUAN site. However, because all maintenance of an instrument can also introduce discontinuities in measurement series, maintenance shall not be conducted more frequently than is necessary. Maintenance schedules must be developed for all instruments. All maintenance actions on instruments shall be documented as part of the metadata associated with the measurements made by that instrument.

Maintenance of supporting infrastructure at GRUAN sites is also essential, particularly in regards to maintaining those aspects of the environment that may affect measurements, such as the painting of Stevenson screens, controlling the growth of trees which may impinge upon the field of view of optical instruments, and maintaining environmentally controlled facilities for those instruments that require it.

7 MEASUREMENT SCHEDULING

7.1 Responsibilities

The WG-GRUAN shall work with the appropriate task team (in the first instance the *Measurement schedules and associated instrument-type requirements* task team, hereafter in this section referred to as 'the task team') to define measurement schedules that allow the resultant data products to capture all important scales of temporal variability to meet the multiple needs of GRUAN data users. These schedules should be conservative in the early stages of GRUAN to allow the appropriate task teams to refine their studies since currently there is a range of opinions on what is necessary. Schedules will depend upon instrument characteristics and expendables and the availability of site personnel, so there cannot be a one-size fits all solution. Therefore, exact measurement scheduling requirements and recommendations will most efficiently form a part of each instrument specific technical document.

When GRUAN operations have been implemented, the core measurement schedules and associated instrument-type requirements shall be agreed upon by the Lead Centre with individual sites, subject to the agreement of WG-GRUAN as part of the assessment and certification process. Subsequent changes to the GRUAN operations at a site must be notified to the Lead Centre and then implemented, as far as possible, by negotiation between the Lead Centre and the GRUAN sites.

Measurement scheduling shall remain stable unless there is a clear requirement for change, which would then have to be agreed with the relevant GRUAN sites. Amendments to the GRUAN measurement scheduling protocol shall be submitted by the task team to the WG-GRUAN before being distributed to GRUAN sites for implementation. Measurement programmes at GRUAN sites are likely to be constrained by more than just the requirements of GRUAN. In recognition of the heterogeneity of the network, the scheduling protocols defined in this document may not apply at every GRUAN site, but any deviation from the measurement schedule must be agreed by the GRUAN Lead centre and then accepted by WG-GRUAN. Individual GRUAN sites will define which measurements and measurement schedule they can sustain as part of the certification and assessment process (Section 5.5).

For designing measurement schedules for process understanding, it will be necessary for the task team to work closely with individual sites since scheduling in support of process understanding is more likely to be site specific. For example, some sites are more likely to experience specific synoptic conditions related to the understanding of associated processes compared to other sites. The primary responsibility of the task team and WG-GRUAN is to determine what understanding of mesoscale processes is required for climate purposes but the primary responsibility for ensuring that measurement schedules lead to such understanding lies with the sites.

Given that task teams have a finite operating life, should the task team no longer exist, this scheduling guidance responsibility shall fall to selected members of the WG-GRUAN who may co-opt participants from the wider GRUAN community to assist with revising measurement scheduling protocols.

7.2 Guiding principles

Where available, scientific and statistical studies shall inform the process for establishing measurement schedules. These shall be published in the peer reviewed literature wherever possible. However, a sound scientific basis for the measurement schedules discussed in this document are not always available and until they become available, the measurement schedules must be considered to be preliminary.

Some evolution of measurement schedules can be expected in the longer term when performance requirements from the network for climate studies become clearer, but such changes must be strictly limited in frequency and agreed to by the WG-GRUAN, and the GRUAN sites.

In cases where oversampling would allow averaging of measurement results to reduce the type-A uncertainties, and where this is technically feasible, measurement schedules should be set accordingly.

The highest priority is that measurement schedules are established to meet the needs of the four primary communities of users of GRUAN data products, viz.: the climate detection and attribution community, the satellite community, the atmospheric process studies community, and the numerical weather prediction (NWP) community. Where perturbations to schedules would increase the utility of the measurements for one of GRUAN's user communities without compromising the needs of another, measurement schedules should be adapted. For example, as long as it does not compromise long-term trend detection, the timing of a daily measurement may be shifted to coincide with a satellite overpass and in this way provide valuable high quality data for satellite validation.

Where possible, measurement schedules for redundant systems should be synchronized so as to avoid introducing biases resulting from differences in timing of the two measurements when combining the measurements into a single data product.

Required measurement schedules may vary regionally and seasonally. In places and seasons where the parameter being measured is more variable, measurements should be made more frequently so that the effects of that variability can be captured.

Where regression models might be used to statistically attribute observed changes in some climate variable to a number of different drivers of those changes, measurement schedules should be set in such a way that the attribution can be conducted in a statistically robust manner. For example, if the effect of some forcing varies diurnally, and that diurnal effect is to be captured, the measurements should sample the full diurnal cycle.

A first step is to ensure that the sampling does not introduce biases into derived monthly means, followed by a second step which determines how those monthly mean uncertainties affect the statistical robustness of trends derived from those monthly means. For example, over Antarctica ozone changes rapidly during the month of October. At high southern latitude sites, ground-based measurements making use of the Sun as a light source (e.g. Dobson spectrophotometers) are often concentrated towards the latter half of October since earlier the solar elevation is too small. A monthly mean calculated from such sampling would be biased. Similar caveats apply for sampling of constituents which show strong diurnal variations (Wang and Zhang, 2008).

Meteorological reanalyses and/or models, such as atmosphere-ocean general circulation models or coupled chemistry-climate models, will have a useful role to play in guiding measurement scheduling. They can be used to provide initial estimates of the autocorrelation, the magnitude of variability, and the trend in climate variables or atmospheric composition as a function of location and season. Simulating the effects of a measurement schedule by sampling reanalyses or model output on the same schedule as the measurements can provide an indication of how the proposed measurement schedule is likely to affect the determination of variability on a range of timescales as well as long-term trends (Seidel and Free, 2006). Where data for a specific atmospheric variable are not available, analysis of temperatures can often serve as a valid proxy for other climate variables since temperature responds to climate variability in a similar way to many other climate variables. It must be recognized, however, that both models and reanalyses may not provide a completely accurate representation of atmospheric means and variability.

For some measurements, scheduling with respect to UTC or LST may be important and may result in conflicting requirements regarding different intended uses of the measurements. For example, scientifically it may be advantageous to have all GRUAN sites making measurements at the same LST (especially for variables that show strong diurnal variations or for instruments that have diurnally dependent biases that must be minimized), while for ensuring coincidence with GUAN sites having all measurements made at the same UTC might be more appropriate.

Current assimilation schemes used in NWP and reanalysis centres, e.g. 4D-var assimilation, are more able to make use of measurements made at different times of day than earlier assimilation schemes. Therefore, consistent timing of measurements is not a prerequisite for assimilation into NWP or for future reanalyses. If, however, the variable being measured shows a strong diurnal cycle, or if the instrument being used to make a measurement has diurnally varying biases, changing measurement times would introduce additional variability which would need to be accounted for in any analysis to avoid sampling bias.

A discussion of frequency of measurements cannot occur outside of a context of a particular anticipated scientific study. The characteristics of a measurement that are deemed sufficient for a particular science study will change depending on what analysis is intended to be done. Thus, in order to specify the required frequency of measurement and its accuracy, precision, temporal and spatial resolution, it is also necessary to specify what analysis will be done with the measurements. Given the intended long-term nature of GRUAN it is impossible to predict, with accuracy, what such future studies may entail so a degree of caution is warranted against over-specification which may later prove deleterious.

7.3 Factors affecting measurement scheduling for trend detection

For trend detection the following factors should guide the development of measurement schedules (Weatherhead et al., 1998):

7.3.1. The magnitude of the variability

In most cases this will vary as a function of season. Where measurements through a month are sparse, and where monthly means of those measurements will be used in trend analysis, the day-to-day variability within the month will determine the representativeness of the sparse measurements in quantifying the true monthly mean. The variability in the monthly means themselves, or, more precisely, the variability in the monthly residuals after a regression model fit, will also determine the statistical robustness of derived trends.

7.3.2. Autocorrelation

This is a measure of the ‘momentum’, ‘memory’ or ‘latency’ in the system. When autocorrelation is high, measurements in consecutive time periods are highly correlated. When autocorrelation is low, consecutive measurements are largely independent of each other. Autocorrelation is also likely to vary through the year. If monthly means of measurements are being used in trend detection, the auto-correlation between those monthly means constitutes an important source of uncertainty in the trend estimate (Tiao et al., 1990) – when autocorrelation is high, the uncertainty on the estimated trend increases. One advantage of using monthly means for the calculation of long-term trends is that the uniform temporal sampling simplifies the calculation of the autocorrelation in the signal. However, individual measurements may also be used in trend detection and methods are available for determining the autocorrelation in such potentially unequally spaced measurement time series (Bodeker et al., 1998). A clear distinction must be made between:

i) Day-to-day autocorrelation which determines, in part, the likelihood of over-sampling,

ii) Day-to-day variability which determines the robustness of the monthly mean when it is calculated from sparse, isolated measurements through the month,

iii) Autocorrelation in the monthly means which determines, in part, the uncertainty on calculated trends,

iv) Variability in the monthly means which also contributes to the uncertainty on calculated trends.

All four of the above are likely to vary spatially and seasonally with the result that optimal measurement schedules are likely to vary between sites and with season. On initiation of a measurement programme, and where the autocorrelation is not known *a priori*, measurements should be made at the highest possible frequency so that a robust seasonal pattern of the autocorrelation can be established. Thereafter, measurement frequency can be relaxed during periods of expected high autocorrelation since momentum in the system being sampled will result in nearby measurements not being independent.

7.3.3. The random uncertainty of the measurement

When measurements with small random uncertainty can be made, measurement frequencies can potentially be reduced, depending on the extent to which random uncertainty is a factor in trend detection or in analyses of specific atmospheric phenomena. When random uncertainties are larger than the expected target uncertainty, high frequency sampling is required to reduce their effects on the total uncertainty of the mean. Random uncertainty and systematic biases can also vary with season as a result of confounding factors (such as surface albedo, humidity and temperature) which vary through the year. The derivation of measurement uncertainties within GRUAN, and how these might vary with season, must therefore play a role in determining measurement scheduling.

7.3.4. The size of the expected trend to be resolved

For large trends, the signal to noise ratio will be high and measurement frequencies can be reduced (all else being equal).

7.3.5. The seasonality in the trend

For many ECVs the magnitude of the trend is likely to be a function of season. Measurement schedules must therefore be set so that statistically robust monthly means, with well characterized uncertainties, can be derived for each month of the year.

7.3.6. Discussion

Where the random uncertainty on the measurement is a significantly smaller contributor to the uncertainty in the trend estimate than autocorrelation and natural variability, for most mid-latitude locations and for most climate variables, the autocorrelation in the system results in diminishing returns for measurements made at a frequency of higher than every 3 days. On the other hand, for most climate variables measured at mid-latitudes, sampling less frequently than every 10 days significantly increases the uncertainty on derived monthly means.

The interplay between the four factors discussed above must be accounted for when planning measurement schedules. It may be that the uncertainty on derived trends is limited by natural variability rather than by the random error on the measurement, in which case more resources should be invested in increasing measurement frequency rather than reducing the random errors. In some cases this may require a cost-benefit analysis where the cost to detect a putative trend of X%/decade (perhaps based on projections from models) over N years is minimized. A cheaper instrument making a more uncertain but more frequent measurement might be selected over a more expensive instrument making a less uncertain but less frequent measurement, since the greater frequency leads to detection of the expected trend either in fewer years or at a lower cost. A measurement strategy might have a greater cost per year than any alternative, but if that strategy can detect a statistically robust trend in fewer years, the net cost may be reduced. However, the detection of statistically robust trends in upper-air ECVs is not the only purpose of GRUAN sites and the cost-benefit analysis for any measurement scheduling protocol must remain cognizant of all intended uses of GRUAN data and the multi-decade measurement programmes expected of GRUAN sites.

7.4 Interplay of science goals and scheduling frequency

As an example of the interplay between science goals and scheduling frequency, this section considers the needs for water vapour measurements within trend detection, satellite validation, and process studies. This example highlights the different issues that need to be considered when developing a measurement schedule to meet a set of scientific objectives at any particular site, and provides some general scheduling guidelines.

7.4.1. Trend detection

In considering the needs for trend detection, two atmospheric regimes of greatly differing characteristics are considered.

Upper troposphere

Recent work (Soden et al., 2005, Boers and Meirgaard, 2009, Whiteman et al., 2011) indicates that the largest anticipated trends in atmospheric water vapour amount may occur in the upper troposphere with increases of up to 1% or more per year in absolute humidity on average over the coming century. The large variability in upper tropospheric water vapour implies that even using the most accurate sensors possible, time to detect trends in the upper troposphere will likely take 20 to 30 years or more depending on what uncertainties in the calculated trend values are tolerable. This large variability, however, also implies that the trend calculations are relatively insensitive to random uncertainty in the water vapour measurements themselves. The greatest decrease in time to detect trends is realized by increasing the measurement frequency as opposed to decreasing the random uncertainty of the measurements. For example, random uncertainties of up to 50% and greater are tolerable for measurements of water vapour in the upper troposphere without significantly impacting the time to detect trends. However, relatively large random uncertainty in a measurement can make the presence of small systematic uncertainty more difficult to both detect and robustly account for. Therefore, if time series are to be developed from instruments with inherently higher random uncertainty in the upper troposphere (e.g. Raman lidar), procedures should be implemented that tend to randomize sources of systematic uncertainty. An example of such a procedure would be frequent re-calibration of the instrument with respect to an external reference. The study of Whiteman et al., 2011 found that measurements acquired approximately twice per week offered perhaps an optimum trade-off between time to detect trend and cost of measurements. The vertical resolution required for upper tropospheric measurements has not been directly studied but a mean value in a layer of 1 km thickness in the upper troposphere would likely have adequate vertical resolution for the purposes of trend detection. The accuracy and stability needs for these measurements have also not been directly studied but if techniques can be devised that tend to merge both of these parameters into the random uncertainty budget, their influence will be decreased.

i) *Systematic uncertainty*: not yet studied but 5-10% is likely adequate. Accuracy becomes less important if recalibrations randomize this component of the uncertainty budget over time.

ii) *Random uncertainty*: up to 50% with the caveat that large random uncertainties can mask small systematic uncertainties

iii) *Stability*: not yet studied although changes in calibration are known to increase the time to detect a trend. Stability becomes less important if procedures randomize this component of the uncertainty budget over time.

iv) *Temporal and spatial resolution*: not yet studied but vertical resolution of 1 km or less is likely adequate. Temporal resolution on the order of an hour or less would appear adequate.

v) *Time of day to sample*: not yet studied, but the lack of a causal connection between upper tropospheric humidity and time of day would imply that day or night sampling or a combination of the two should be equally effective at revealing trends.

<u>Lower stratosphere</u>

Detailed studies of the time to detect water vapour trends in the lower stratosphere have yet to be completed. The modelling work that has been done indicates that anticipated trends in the lower stratosphere can be expected to be smaller than in the upper troposphere although stratospheric modelling may have larger uncertainties associated with it than comparable work in the upper troposphere. Despite this relative lack of knowledge, certain general statements can still be made regarding measurement needs in the lower stratosphere. First, lower stratospheric water vapour variability is dramatically lower than in the upper troposphere. This almost certainty implies that the calculations of trends in the lower stratosphere will require measurements of much higher accuracy and precision than in the upper troposphere. We expect that, even with measurements with small uncertainties, increased measurement frequency will still be desired to decrease the time to detect trends although specific guidelines for measurement frequency in the lower stratosphere are not yet available. Regarding vertical resolution, recent work (Hurst et al., 2011b) shows that trends in vertical layers of 1 to 2 km thickness need to be resolved so the measurement systems should provide high accuracy measurements in layers of approximately 1 to 2 km in thickness if these 'sub-trends' are to be revealed. Given that measurements with small uncertainties are likely required to reveal trends in the lower stratosphere, high stability is likely also required. The same recommendation applies as in the upper troposphere – procedures that tend to randomize sources of systematic uncertainty will create a higher quality data set over time.

i) *Systematic uncertainty*: not yet studied, but detecting trends in the lower stratosphere will be much more sensitive to sources of systematic uncertainties than in the upper troposphere. However, practical issues currently limit the potential performance to 5-10% calibration uncertainty. Instrument developments to improve this would be valuable, and procedures that randomize this uncertainty in the long-term would also be beneficial.

ii) *Random uncertainty*: not yet studied, but detecting trends in the lower stratosphere will be much more sensitive to sources of random uncertainties than in the upper troposphere. The recommendation is therefore a 10% random uncertainty maximum.

iii) *Stability*: not yet studied, but detecting trends in the lower stratosphere will be much more sensitive to changes in calibration and other non-random errors that contribute to measurement instability over the longer term. So procedures that randomize these uncertainties are beneficial.

iv) *Temporal and spatial resolution*: The recommendation is 1 km vertical resolution or less in order to reveal sub-trends as discussed in Hurst et al., 2011b.

v) *Time of day to sample*: not yet studied, but the lack of a causal connection between lower stratospheric humidity and time of day would imply that day or night sampling or a combination of the two should be equally effective at revealing trends.

7.4.2. Satellite validation and radiation studies

The discussion concerning the measurement needs for the purposes of satellite validation will be broken into the needs for comparisons to be done either in radiance space or in retrieval space. The discussion on radiance space comparisons will also discuss uncertainties in determining outgoing longwave radiation (OLR) since these also provide some guidance for satellite radiance validation studies.

Radiance comparisons using a forward model and considerations of OLR errors

The brightness temperatures measured by passive space-borne sensors are calibrated with small systematic uncertainty. For example, the Atmospheric Emitted Radiance sensor frequency-dependent brightness temperature uncertainties (V3.0 validation report) were specified to range from 0.1 to 0.5 K with systematic uncertainties typically much less than 0.1 K. Considering the upper troposphere and using the rule of thumb from Soden et. al, 2000 that a 1 K difference in brightness temperature corresponds to a change in upper tropospheric water vapour amount of about 12%, the systematic uncertainties in the AIRS radiances, themselves, translate into negligible uncertainties in upper tropospheric water vapour amounts. However, to quantify the water vapour amount from brightness temperature requires the use of a forward model which may have substantially larger uncertainties in spectroscopy. Past efforts have shown that absolute accuracy of water vapour measurements in the upper troposphere of approximately 5% were sufficient to reveal small spectroscopic uncertainties in forward model studies. Given that 5% systematic bias in the upper troposphere water vapour measurements is unlikely to be achieved with current technology on a routines basis, this is an area where technology improvement can have significant impact. It is possible that only specially processed datasets from campaign mode periods will possess the accuracy required for this type of stringent study.

It is also useful to consider the data requirements for radiation closure studies within the context of the satellite validation topic since this area of research is something that a well-characterized column will permit and the measurement needs are in some ways similar to those for satellite radiance validation. Ferrare et al., 2004 consider the OLR consequences of errors in water vapour concentration as a function of altitude. They show, for example, that a 3% uncertainty in total column water vapour amount results in an uncertainty of 0.5 W/m^2 in the outgoing longwave radiation. A 10% uncertainty in the upper 0.1 mm of total column water (typical column amounts in the upper troposphere) results in the same uncertainty in OLR. We can take these numbers then as additional guidelines for accuracy needs for total column water and upper tropospheric water vapour measurements.

i) *Systematic uncertainty*: total column water vapour amount 3%, 5% profile systematic uncertainty in the lower and middle troposphere, 10% in the upper troposphere.

ii) *Random uncertainty*: measurement needs depend on the statistics of the investigation being done. If there are many comparisons, relatively large random uncertainties are tolerable (guideline of 10 – 20%). If individual comparison case studies are attempted, random uncertainties must be low (guideline <= 5%)

iii) *Stability*: not explicitly studied but if studies are done over a short period of time, most of the concern regarding data quality relates to the random and systematic measurement uncertainties.

iv) *Temporal and spatial resolution*: given that passive satellites measure typically in a fraction of a second for a given scene, high temporal resolution is desirable. Where this is not feasible, comparisons made under conditions of low atmospheric variability are desired. Data handling procedures that reduce the influence of atmospheric variability on the processed results are desired.

v) *Time of day to sample:* at the time of the satellite overpass. A radiosonde launch should precede the actual overpass so that the sonde is approximately in the mid-troposphere at the time of the overpass. Lidar measurements that are averaged over time can make use of variable temporal integration as a function of altitude. Knowledge of the local atmospheric variability would enable the additional uncertainty introduced by the spatial and temporal separation between the measurement and the satellite footprint.

Satellite comparisons in retrieval space

Guidelines for water vapour measurement needs for the validation of hyper-spectral sounders such as AIRS can be obtained from the table of validation requirements and goals for the instrument. Here the desire for AIRS retrievals was 5% random error in total column water and 15% random error in 2 km thick layers. Taking this as a guideline for retrieval validation yields the following guidelines for water vapour measurements intended for satellite retrieval comparisons.

i) *Systematic uncertainty*: 3% total column, 10% in 2-km layers.

ii) *Random uncertainty*: measurement needs depend on the statistics of the investigation being done. If there are many comparisons, relatively large random uncertainties are tolerable (guideline of 10 – 20%). If individual comparison case studies are attempted, random uncertainties must be low (guideline <= 5%)

iii) *Stability*: not explicitly studied but if studies are done over a short period of time, most of the concern regarding data quality relates to the systematic and random measurement uncertainties.

iv) *Temporal and spatial resolution*: given that passive satellites measure typically in a fraction of a second for a given scene, high temporal resolution is desirable. Where this is not feasible, comparisons made under conditions of low atmospheric variability are desired. Data handling procedures that reduce the influence of atmospheric variability on the processed results are desired.

v) *Time of day to sample*: at the time of the satellite overpass. A radiosonde launch should precede the actual overpass so that the sonde is approximately in the mid-troposphere at the time of the overpass. Lidar measurements that are averaged over time can make use of variable temporal integration as a function of altitude. Knowledge of the local atmospheric variability would enable the additional uncertainty introduced by the spatial and temporal separation between the measurement and the satellite footprint.

7.4.3. Process studies

Investigations of various atmospheric phenomena such as frontal passages, drylines and convection initiation have been performed using data from balloon-borne and ground-based remote sensing instruments (Melfi et al., 1989; Demoz et al. 2006; Koch et al., 2008; Bennett et al., 2010). Revealing the details of these atmospheric phenomena is supported by high frequency radiosonde launching but the fine details are missed even with the most frequent radiosonde sampling schedules. Remote sensing systems such as lidars and radars have been found to be a great aid to understanding in these kinds of process studies. Of most benefit are those instruments that provide measurements of key atmospheric parameters (e.g. boundary layer height, aerosol and cloud structures, winds, water vapour content) at high temporal and spatial resolution. Ideally these high resolution measurements start at the ground, extend at least to the mid-troposphere and are available continuously during periods of peak interest. The challenges to be met with such a remote sensing system for GRUAN is for it to demonstrate the capability of making reference quality measurements, provide useful data from the ground upwards, be able to operate day and night and have sufficient sensitivity to be able to probe the convective boundary layer with good statistics in 5 minutes or less throughout the day and night.

i) *Systematic uncertainty*: small systematic uncertainties are not necessarily needed to support process studies. Often it is variations in the water vapour that are most important. Therefore a guideline of 10% systematic ucnertainty is likely adequate.

ii) *Random uncertainty*: in general, process studies are not areas where multiple comparisons can be accumulated to improve the statistics. It is more likely that each case being studied is unique. Therefore random uncertainty requirements need to be more stringent, but the type-A target uncertainty will depend on the exact process under study. General guidelines might be less than 10-25% but random uncertainty requirements will more likely need to be set by the individual investigators based on their individual needs.

iii) *Stability*: process studies are generally short term in nature and stability of measurement systems should not be a large concern.

iv) *Temporal and spatial resolution*: high temporal and spatial resolution is useful. This is an area of particular strength for active remote sensing systems such as radar and lidar. For the case of water vapour, the most highly variable atmospheric state parameter, temporal and spatial resolution of approximately one minute and 10 m – 100 m is desirable. Frequent and, if possible as in the case of an automated system, continuous measurements are desired.

v) *Time of day to sample*: Before and during the event of interest. To be determined by the scientific goals of the experiment, but day and night-time measurement capability is desired.

7.5 Instrument specific measurement schedules

Ideally an assessment, as presented for water vapour in Section 7.4, would be available for each of the ECVs targeted by GRUAN. However, these are not yet available. This section provides some interim instrument specific measurement schedules that can guide operations at GRUAN sites until a sound scientific assessment has been developed for each ECV. Instrument specific protocols will be developed over time and, in the case of conflict between this guide and an instrument specific technical document, the schedule outlined in the technical document takes precedence.

7.5.1. Generic measurement schedules

Once a site has selected the frequency with which measurements will be made, this section provides guidelines on appropriate timing of those measurements. The frequency of measurements at sites will determine, in part, the added value that a site brings to the network (see Section 5.4). This section defines a set of generic measurement schedules which can then be applied and adapted in various circumstances.

<u>Schedule A</u>

This schedule is designed for instruments making one or more measurements per week. Where there is an expected seasonal cycle in natural variability and that cycle is not yet known, intervals between measurements should be constrained by $(4/N) < t < (10/N)$ where t is the interval in days and N is the number of measurements being made each week. Under such a schedule, on average, $52 \times N$ measurements will be made each year. Once a climatology of the seasonal cycle in natural variability has been determined, during the 5 months of the year exhibiting highest natural variability, intervals between measurements should be constrained by $(3.5/N) < t < (6.5/N)$ where t is the interval in days; this should result in a total of $N \times 30$ measurements through those 5 months. For the remaining 7 months of the year intervals between measurements should be constrained by $(7/N) < t < (13/N)$; this should result in a total of $N \times 22$ flights through those 7 months. On average, this will result in $52 \times N$ measurements being made each year but with a higher frequency ($\sim N \times 6$/month) in the months of higher natural variability and a lower frequency ($\sim 3 \times N$/month) during months of lower natural variability. Within the measurement windows defined above, measurement times should be selected to maximize the altitude reached based on knowledge of local conditions, to maximize coincidence with relevant satellite overpasses, and to minimize factors that may contribute to measurement uncertainty e.g. making flights at night rather than during day for instruments requiring corrections for solar heating.

<u>Schedule B</u>

This schedule is designed for instruments making one or more measurements per month. Where there is an expected seasonal cycle in natural variability and that seasonality is not yet known, intervals between measurements should be constrained by $(20/N) < t < (40/N)$ where t is the interval in days and N is the number of measurements being made each month. Once a climatology of the seasonal cycle in natural variability has been determined, during the 4 months of the year exhibiting highest natural variability, intervals between flights should be constrained by $(15/N) < t < (25/N)$ where t is the interval in days; this should result in a total of $N\times6$ flights through those 4 months. For the remaining 8 months of the year, intervals between flights should be constrained by $(35/N) < t < (45/N)$; this should result in a total of $N\times6$ flights through those 8 months. As with Schedule A, within the measurement windows defined above, measurement times should be selected to maximize the altitude reached based on knowledge of local conditions, to maximize coincidence with relevant satellite overpasses, and to minimize factors that may contribute to measurement uncertainty e.g. making flights at night rather than during day for instruments requiring corrections for solar heating.

7.5.2. Radiosondes

Noting the minimum entry level requirements for a GRUAN site (Section 5.2.2):

i) *For sites performing four radiosonde flights daily:* As for a fully equipped GRUAN site (see Section 5.2.1).

ii) *For sites performing twice daily radiosonde flights:* One flight at 00 LST and one flight between 06 LST and 18 LST timed to maximize coincidence with any satellite overpass measuring the same variables or with a redundant measurement made by another instrument at the site. Since satellite-based measurements are more likely to be daytime measurements, the daytime radiosonde launch time is the one which is varied.

iii) *For sites performing daily radiosonde flights:* One flight at 00 LST.

iv) *For sites performing weekly radiosonde flights:* Nominal launch times should be 00 LST on the same day of the week, but allowed to vary by up to 48 hours either side to match satellite overpasses or to match the timing of redundant measurements.

v) *For sites performing monthly radiosonde flights:* Nominal launch times should be 00 LST on the same day of the month, but allowed to vary by up to 5 days either side to match satellite overpasses. It would be expected that these would be high quality sondes and launch times should also be selected so that conditions most likely to lead to measurements as high in altitude as possible are achieved.

7.5.3. Frost point hygrometers, ozonesondes and aerosol sondes

i) *For sites making one or more flights per week*: Schedule A (see Section 7.5.1)

ii) *For sites making one or more flights per month*: Schedule B (see Section 7.5.1)

7.5.4. GPS integrated precipitable water

The GNSS receivers at GRUAN sites shall track GNSS satellites with a sampling interval of 30 seconds or less. The minimum requirement for GNSS raw data submission is daily (24 hour) files with a 30 second sampling interval.

Surface meteorological observations shall be made at GNSS sites at intervals of no more than 60 minutes. An observation interval of 10 minutes is preferred.

An hourly sampling interval is required for GNSS tropospheric products and associated supplemental data, including zenith tropospheric delay, zenith wet delay, precipitable water, surface pressure and atmospheric water-vapour-weighted mean temperature.

7.5.5. Raman lidars

Continuous measurements 24 hours a day, 7 days a week is technically possible for lidar. In practice, and considering the instrumental and human constraints, only a limited number of lidar systems can achieve sustainable continuous 24/7 operations. Lidars can measure in clear sky as well as thin clouds. Logistical and financial support permitting, GRUAN lidar instruments having a 24/7 capability should adopt the 24/7 schedule as their default schedule. When logistical and/or financial support precludes 24/7 operation, default schedules must be chosen to address one or several of the following considerations: long-term variability studies, process studies, satellite validation, and GRUAN measurement redundancy. A minimum of 6 hours per week spread over 2 to 4 nights of operation may be suitable for long term monitoring. Additional details can be found in Section 3.1 of the GRUAN Lidar Guidelines document, which applies the general scheduling guidelines described in Section 7.5 specifically to lidar operations.

When redundancy between programmes at the same GRUAN site can be identified, the lidar should be operated according to the following recommendations:

i) *For sites performing at least daily radiosonde flights:* the lidar does not need to be operated every night, but when operated, its running time should be coincident with the first night-time flight of the day. The first half-hour of the radiosonde flight must fully encompass the lidar data acquisition period, i.e., must be included between lidar start and end times.

ii) *For sites performing weekly or monthly radiosonde flights:* the lidar must be operated at least on the nights (days) of the radiosonde flights. The first half-hour of the radiosonde flight must fully encompass the lidar data acquisition period, i.e., must be included between lidar start and end times

iii) *For sites performing frost-point hygrometer (FPH) flights:* the lidar must operate at least on the nights (days) of the FPH flights. Extended hours of lidar operation (e.g. all night or at least 4-5 hours) are recommended in an attempt to extend and/or optimize the profiles in the UT/LS. The first full hour of the FP flight must fully encompass the lidar data acquisition period, i.e., must be included between lidar start and end times.

7.5.6. Microwave radiometers

Off-the-shelf commercial microwave radiometers are robust and unattended instruments that provide real time accurate atmospheric observations 24 hours a day, 7 days a week. These units can perform under all-weather conditions, though the quality of retrieved atmospheric parameters degrades under conditions of precipitation. The level of degradation depends on precipitation intensity and any mitigation solutions adopted, including rain sensor, hydrophobic coating, tangent blower, shutter, and side-view, should be recorded.

Accurate observations are subject to instrument integrity and proper signal calibration. Commercial units consist of robust hardware exhibiting long life-time (years) even in extreme conditions. However, the dome protecting the antenna aperture must be kept clean, requiring regular servicing and replacement every few months, depending on environmental conditions (presence of dirt, sand, dust, etc.). The current technology is such that calibration is stable over long periods (months). To avoid long periods of mis-calibration, an operational protocol (including strict quality criteria and a testing period) shall be adopted before accepting the calibration coefficient updates.

Commercial units may be equipped with azimuth- and elevation-angle scanning capabilities. Elevation scanning is useful for increasing the vertical resolution of temperature profiles in the planetary boundary layer. When both azimuth and elevation scanning are available, hemispheric observations of IWV, ILW, and temperature can be performed, at the expense of the time observing in the zenith direction.

7.5.7. Microwave spectroradiometers

Microwave radiometers equipped with a spectrometer to spectrally resolve the line shape of the emission line of water vapour are operated at a number of NDACC sites. Such instruments are operated continuously and are normally controlled remotely. Because observations during precipitation are not valid, automated rain gauges are a useful complement to such measurements providing a valuable metadata element.

To achieve a reasonable signal to noise ratio for the measured spectrum, an integration of individual spectra must be performed. The integration time is typically a few hours depending on instrument parameters and atmospheric opacity. Integration times can be shortened under conditions of lower water vapour content of the troposphere. For this reason observations in humid environments, especially in summer, tend to have a poorer temporal resolution than in very dry Arctic conditions. Under optimum conditions a time resolution as good as two hours can be achieved whereas under less favourable conditions daily profiles are realistic for the stratosphere. It has been shown that such instruments can retrieve water vapour profiles down to ~25 km altitude under very dry conditions whereas 30 to 35 km is the more typical lower boundary for water vapour profiles measured by such instruments.

7.5.8. Fourier Transform Spectrometers

Ground-based Fourier-Transform Spectrometer measurements need a clear field of view of the solar disc. Measurements cannot be performed under complete to moderately cloudy conditions. In the event of e.g. thin high cirrus, measurements can be made but the signal-to-noise ratio (SNR) and consequent measurement uncertainty will suffer. The infrared region is covered by six or more different spectral filter regions, which assures an optimal SNR. The measurement of each filter region takes 2 to 10 minutes. This measurement integration time is a function of the spectral resolution and the required SNR. On a clear day, spectra can be recorded continuously for solar elevations above about 5° (at lower elevations the uncertainty might increase). For an instrument dedicated to measuring only water, measurements could be made approximately every 3 to 5 minutes. Conversely a more versatile configuration observing the entire mid-infrared would make a repeatable series of measurements in approximately one hour. Due to the clear sky constraint, routine or regularly scheduled measurements are not strictly made. A typical automatic system might attempt observations daily, taking the opportunities that weather conditions provide. This could be one to many per day.

The instrumental line shape is calculated every few months by measuring the transmittance spectrum of a standard low pressure cell. This calibration, which provides a metric of instrument performance, can be performed automatically or manually. The infrared detectors require liquid nitrogen daily.

7.6 Operation and maintenance, calibration andquality standards

Standards of operation, calibration and maintenance for each instrument used in GRUAN are developed and detailed in instrument specific technical documents to ensure that minimum quality standards are achieved and maintained. This will be necessary to minimize sources of uncertainty when measurements are being made using sophisticated instruments that may not always be completely familiar to the operator. This will be more likely the case when measurements are being made under operational conditions. Operation, calibration and maintenance protocols are such that collection of detailed metadata is mandatory as these metadata will be vital in establishing measurement uncertainties. These metadata may include ancillary observations such as precipitation, cloud cover, temperature at the site or humidity at the site.

8 DATA MANAGEMENT

8.1 Overview of GRUAN data flow

A schematic representation of the flow of data within GRUAN and from GRUAN to the user community is shown in Figure 3.

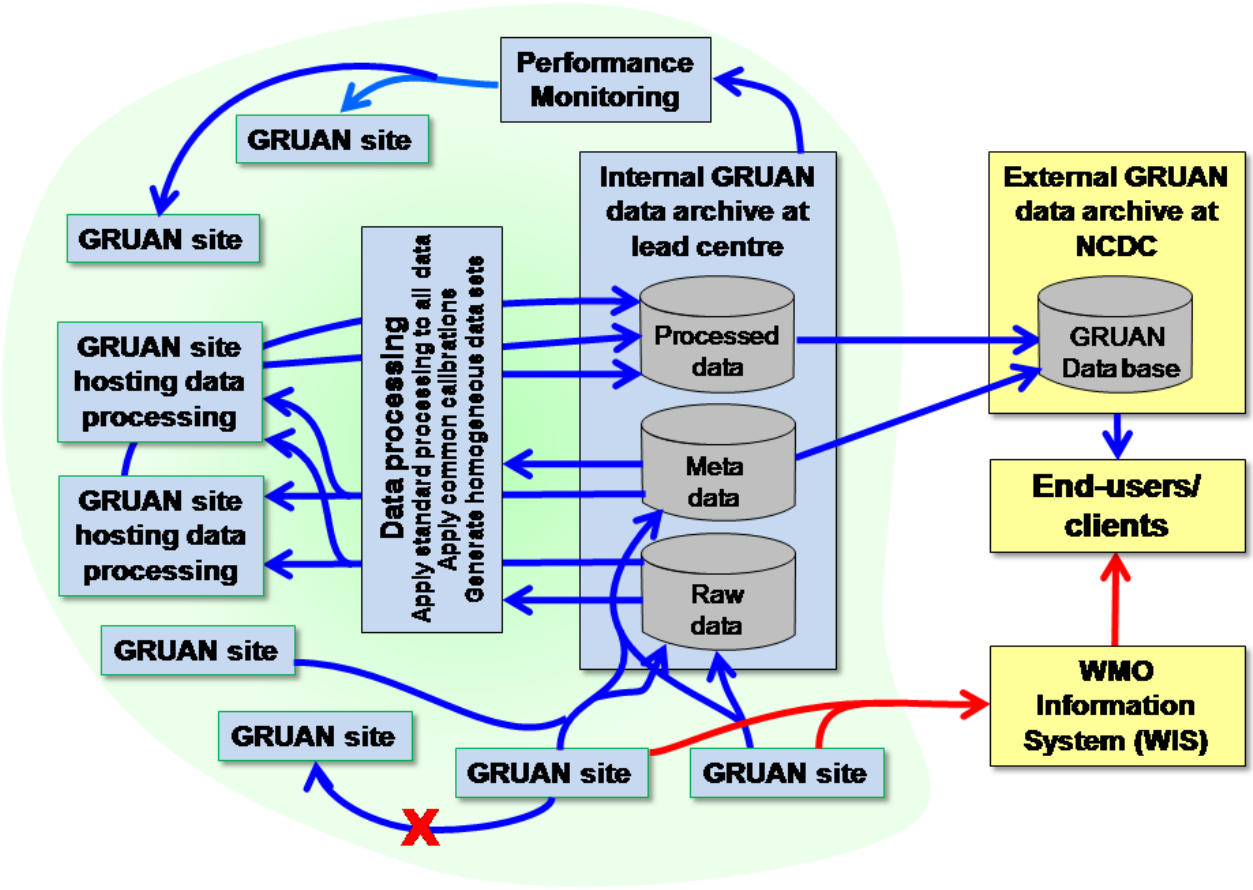


Figure 3: A schematic representation of the flow of data in GRUAN. Blue arrows show the standard flow of data. The red arrows show the flow of near-real time data. Data provided to end-users via red routes are not 'GRUAN data'. Central processing is conducted either at the Lead Centre, at a GRUAN site specialising in the processing of that data stream, or at a third party processing centre. Different data exchange protocols should operate for exchange of data within GRUAN (shaded green region) and from the GRUAN external data archive to end-users. Data originate at GRUAN sites and flow as raw data to centralized processing facilities before returning to the GRUAN data archive at the Lead Centre. Final GRUAN data products are served to end-users/clients through the external GRUAN data archive currently at NCDC.

To avoid ambiguity with numbered satellite data levels, this Guide uses the following nomenclature for GRUAN data levels:

Primary Raw Data (PRD): This is the 'rawest' form of data available e.g. measured voltages before any processing has been applied. Even for the same instrument, formats of PRD data files are likely to differ between sites. PRD are expected to be archived in perpetuity at the site where the measurements took place, at the internal GRUAN data archive at the Lead Centre, and at the nominated GRUAN central data processing facility for that product.

Converted Raw Data (CRD): These data are stored in a common well-described file format intended for long-term storage. They are pre-processed raw data and might already represent parameters to be used in end-user's application, e.g. brightness temperature for microwaves or zenith total delay for GPS. CRD are expected to be stored at the site where the measurements took place, at the internal GRUAN data archive at the Lead Centre, and at the nominated GRUAN central data processing facility for that product.

Near-real-time Data (NRTD): This is a data product resulting from preliminary processing of the GRUAN data subject to as many of the additional GRUAN processing steps as can be achieved in the nominal 2 hour NRT window (Figure 4). Increasing efficiencies and streamlining of data processing with time is expected to lead to more of the additional GRUAN processing steps being incorporated into the NRTD. When NRTD are submitted on the WIS to analysis centres, they must be flagged as having originated at a GRUAN site so that they can be treated appropriately. NRTD are expected to be stored at the site where the measurements took place, at the nominated GRUAN central data processing facility for that product, at the internal GRUAN data archive at the Lead Centre, and at the analysis centres to which the data are submitted.

Standard GRUAN Product Data (SGPD): The GRUAN product resulting from all processing steps associated with a single instrument. SGPD are expected to be stored at the nominated GRUAN central data processing facility for that product, at the internal GRUAN data archive at the Lead Centre, and currently at NCDC.

Integrated GRUAN Product Data (IGPD): This is a product that results from the combination of measurements from multiple instruments e.g. a SASBE product (Tobin et al., 2006). IGPD are expected to be stored at the nominated GRUAN central data processing facility for that product at the internal GRUAN data archive at the Lead Centre, and currently at NCDC.

A technical document associated with each instrument will define what data constitutes each of these levels.

Measurements and metadata are bound together in each of these data levels. PRD are ingested from all GRUAN sites into the internal GRUAN data archive hosted at the Lead Centre (see Section 8.5). Direct exchange of PRD between sites is discouraged since this circumvents the data versioning protocols and reduction of the raw data to a common CRD file format. Similarly, direct exchange of CRD between sites is discouraged since this circumvents network wide application of calibration techniques, and other algorithms applied to convert PRD to CRD that would be implemented either at the Lead Centre or at a centralized GRUAN data processing facility (see below).

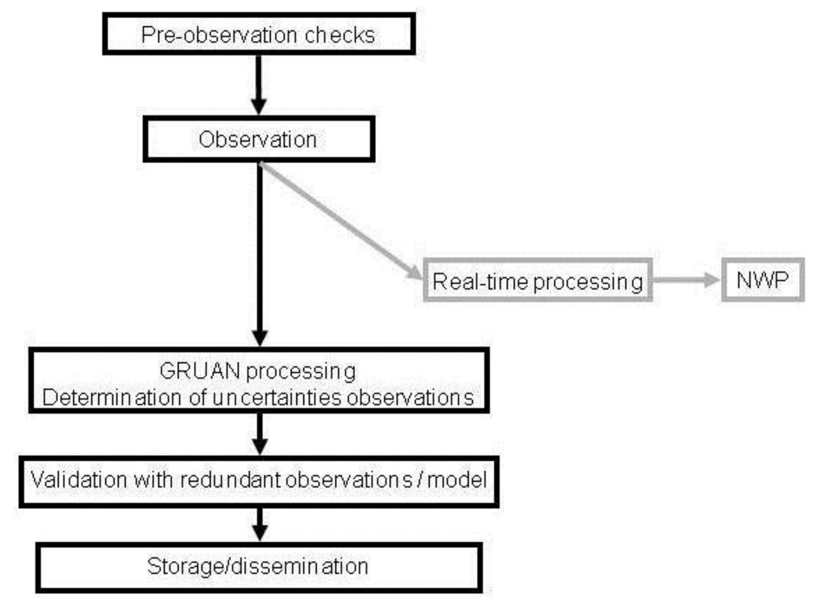


Figure 4: Schematic of NRTD production within GRUAN.

Where GRUAN sites have agreed to the NRT release of their data, these data will be made immediately available via the WIS. This will require some local site-based processing to create NRTD suitable for submission to the WIS.

Processing of the CRD held in the GRUAN internal data archive to produce SGPD and IGPD will occur either at the Lead Centre, at a GRUAN site, or at an internationally recognized facility that specializes in processing data for a particular instrument. This processing would include applying the necessary recalibrations, corrections, and the uncertainty analysis in a consistent and traceable manner across identical instruments from different sites. The SGPD and IGPD, including their metadata and documentation, are provided to the user community through the external GRUAN data archive currently hosted at NCDC. A performance monitoring process (see Section 9), implemented at the Lead Centre, will provide feedback on performance to individual sites.

8.2 GRUAN data policy

This section summarizes and expands on the GRUAN data policy document prepared by the GCOS Secretariat[10]. Since GRUAN is co-sponsored by WMO it is appropriate that any policy for release and dissemination of GRUAN data complies with WMO policy, practice and guidelines for the exchange of meteorological and related data and products. Specifically GRUAN data dissemination and use should comply with WMO Resolution 40 (Cg-XII) which calls for free and unrestricted international exchange of meteorological data and related data and products. Because most GRUAN measurements are considered 'essential' in the context of Resolution 40, they are required to be exchanged without charge and with no conditions on their use. GRUAN sites are likely to provide data to other networks which may have policies in place to protect the rights of the data providers to their own data. No conflict arises here since the data being provided through other networks are not 'GRUAN' data and are therefore not subject to the requirements of Resolution 40.

Three levels of exchange of GRUAN data should be recognised:

i) Exchange of data within the GRUAN community. This should always occur through the GRUAN Lead Centre and/or central data processing centres so that the exchange can be controlled by data policies developed specifically for internal exchange of GRUAN data.

ii) Dissemination of GRUAN products to end-users. This should always occur through the official GRUAN data centre (see Section 8.6). A different policy should be implemented to control the dissemination of GRUAN data at this level.

iii) Dissemination of NRTD on the WMO GTS (Global Telecommunication System)/WIS for assimilation in NWP simulations which occurs via the WMO GTS/WIS rather than through the Lead Centre.

A distinction should be made between 'standard data' and 'enhanced or experimental data' obtained at GRUAN sites:

- Standard data (e.g., near surface synoptic observations, radiosonde observations) have general exploitation value and common measurement technology are generally well understood, and have few problems with data interpretation.
- Enhanced or experimental data (e.g., Raman lidar, microwave radiometer, surface radiation, GPS precipitable water) have high exploitation value, sophisticated measurement technology and/or of experimental nature. Considerable efforts are required to maintain continuous measurements and high quality of the data. Users of these data are recommended to contact site scientist for correct interpretation of data. .

Enhanced or experimental data are more likely to be subject to limitations on dissemination than standard data.

The primary goals of GRUAN (see Section 1.2) are not consistent with NRT dissemination of many of the measurements made at GRUAN sites and certainly not as GRUAN processed meas-

[10] Available from http://www.dwd.de/bvbw/generator/DWDWWW/Content/Projekte/Gruan/Downloads/GRUAN__LC/gruan__data__policy,templateId=raw,property=publicationFile.pdf/gruan_data_policy.pdf

urements. Generating high quality measurements with well characterized uncertainties takes a significant investment of time and effort. In GRUAN the emphasis is clearly on providing reference quality measurements rather than providing NRT measurements. However, it is recognized that measurements at GRUAN sites are likely to be very useful to a number of users requiring data in NRT e.g. for initializing NWP models. Therefore, where possible, and where it does not detract from achieving the primary goals of GRUAN, GRUAN sites should submit NRTD to end-users via the GTS/WIS. The measurements for which NRT submission may be valuable are also more likely to be 'standard data' as described above. The WIS requirements, e.g. on metadata, and the transmission of NRT data via the GTS is strongly encouraged but is not considered a mandatory requirement for GRUAN sites (see Section 5.3). This decision to exclude NRT submission of GRUAN data from the list of mandatory requirements for a GRUAN site is consistent with the recommendation of the AOPC who at their XIVth session stated in recommendation #29 'AOPC recommended that GRUAN data policy should request sites to provide all data in a free and unrestricted manner (in accordance with WMO Resolution 40 (Cg-XII)), and if possible in real time, in order to be of maximum value for all applications'. Where sites do not currently have the infrastructure or expertise to make such submissions, assistance from WMO should be obtained in the form of hardware and/or training. There may be advantages to submitting data in NRT since data assimilation algorithms are able to flag data that appear to be statistically anomalous. If such two way communication can be established between GRUAN and the NWP/data assimilation community, such information could form an important part of the measurement metadata (Section 9). Submission of NRTD will also facilitate the quality control link between GRUAN and GUAN, which will be facilitated by WMO (GCOS-155).

When GRUAN data are used in a scientific publication, the origin of the data must be acknowledged and referenced. A minimum requirement is to reference GRUAN as a reference network of GCOS and to acknowledge the GRUAN data archive currently at NCDC as the source. If data from only one GRUAN site (or a limited number of sites) have been used, additional acknowledgement of those site(s) and their sponsoring institutions or organizations must be given, as specified in the metadata associated with the data files.

Inclusion of GRUAN scientists as co-authors on papers making extensive use of GRUAN data (and in particular enhanced or experimental data) is justifiable and highly recommended, in particular if a site scientist has responded to questions raised about data quality and/or suitability for the specific study in question, or has been directly involved in contributing to the paper in other ways. Co-authorship should not be a pre-condition for release of GRUAN data. However, for enhanced or experimental data it is highly recommended that data users invite site scientists to become co-authors on resultant publications, or determine whether an acknowledgement would be sufficient. Users of enhanced or experimental GRUAN data should be encouraged to establish direct contact with site scientists for the purpose of complete interpretation and analysis of data for publication purposes. GRUAN metadata should include all information related to acknowledgements and/or co-authorship on publications making use of the data. The Lead Centre and Task Team of Site Representatives can facilitate such communication.

8.3 Collation of metadata

Two different types of metadata need to be accommodated within the GRUAN data management facilities, viz.:

i) Metadata describing the context in which the measurement was made i.e. the calibration procedures, data processing algorithms employed, traceability to SI standards, log books, site photos etc. This information will be relevant to a set of data and not specific to any particular datum.

ii) Metadata associated with each datum. For example, for point source measurements, as opposed to column or partial column measurements, in addition to the measurement uncertainty associated with that datum, metadata such as the exact date and time associated with the datum, as well as the altitude, latitude and longitude must be directly available or easily derivable from other metadata. The provision of such metadata recognises the fact that e.g. balloon-borne instruments drift in latitude and longitude during a flight. These data can only be used in 4D-Var assimilation if they are tagged with their 4D (time, latitude, longitude, altitude) coordinates.

All changes to the site, such as site exposure, instrument changes including height above ground, calibrations, inspection visits, data adjustments, and quality control applications are all essential for proper scientific decisions and judgments related to the use of the resultant measurements. Metadata should not preclude information derived from historical documents such as observing practices manuals, site inspection reports, government policies, resource and funding programmes, even local newspapers.

Management and maintenance of metadata requires the investment of resources. Present day technology for database warehousing of digitized metadata has the added benefit that metadata can be accessed, linked to measurements, and easily transferred. To facilitate metadata collation, applications to directly ingest or derive as much metadata as possible from routine operations, such as site inspections, into the GRUAN database need to be developed. Network-wide observation policies and practices, processing algorithms, quality control procedures, data adjustments, units, data formats, etc. should also be maintained to supplement the database management system. Documents related to historical operations at GRUAN sites and to historical data archives should be inventoried and properly conserved until such time as their information content can be transferred to a medium which supports multiple users' access.

Metadata needs to have the same level of commitment as observed data. Incomplete, outdated, or inaccurate metadata can be as detrimental, indeed in some cases worse, than no metadata at all. Regular reviews of metadata content for confirmation and accuracy should be part of regular GRUAN operations. Support to investigate new metadata sources, information management technologies and information sharing capabilities should be on-going in an effort to make accessible and preserve the historical investment in the data collected.

8.4 Data format

In the same way that a distinction should be made between the distribution of data within the GRUAN community and the dissemination of GRUAN data to end-users, a distinction should be made with regard to prescribed data formats for these two different aspects of data distribution, viz.:

i) For distribution of data within GRUAN the emphasis should be on expediency. Different data formats for different instruments should be permitted and not discouraged. Whatever format facilitates quick and automated processing of data and its associated metadata should be used.

ii) For dissemination of GRUAN data to clients, a format should be selected that is flexible enough to allow a common format across all GRUAN products, should have an existing large user-base in the client community, should easily allow the inclusion of metadata in each data file, should be an open format/standard that requires no licensing, should be self-describing, and should have a large suite of readily available tools for manipulating the data files. Perhaps the most suitable format would be NetCDF and better still CF (Climate and Forecast) compliant NetCDF. Tools such as NCO[11] (NetCDF operator) should then be made available for manipulating these files.

8.5 Data submission

If sites elect to submit NRTD to end-users, this should be done directly through the WIS or through their own portals, without a GRUAN label attached but designated as having originated at a GRUAN site. Otherwise all data from GRUAN sites should flow through the Lead Centre. The expectation might be that GRUAN sites submit their raw data to the GRUAN Lead Centre and/or central data processing centres as soon as possible after the measurement but with the policy in place that these data will not be made available outside of the GRUAN community at that time. A facility for imposing time limits on making the data available to the end-user community for different sites should be implemented as this does not contravene WMO Resolution 40 (Cg-XII). In this way sites are more likely to be willing to make their raw data immediately available within the GRUAN community without compromising their rights to first publication of the data (some funding agencies may even insist that such a data policy is in place).

Procedures for submitting data and metadata from GRUAN sites to the GRUAN archive should be developed in such a way as to minimize the effort required at the GRUAN sites and to harmonize the process of data collection and data quality control across the network as a whole. For example, submission of data to the GRUAN archives can be easily automated if the mode of submission is through FTP to a server based at the Lead Centre, whereas if submission must be done through a web portal this cannot be easily automated and is likely to be very time consuming for individual GRUAN sites.

Where data submission tools can be developed centrally (e.g. at the Lead Centre) and distributed for use to GRUAN sites to facilitate data submission to the GRUAN archives, this is preferable to each site independently developing such tools. The ability for sites to jointly contribute to supporting such network wide activities would be desirable.

8.6 Data dissemination

Dissemination of GRUAN data products to end-users/customers shall occur through an official GRUAN data centre currently hosted at NCDC. Access to GRUAN data through a single source will reinforce the model that GRUAN data are homogeneous both in time and across GRUAN sites.

For climate research in particular it is important that users of climate data can, if required, obtain complete information on how the data they are using were acquired. Therefore, users of GRUAN data shall have access not only to the measurements and their uncertainties, but also to the instrument, operating procedures, data reduction algorithms used, and to when changes to any of these occurred through the complete time period of the data set.

A facility should be established whereby users of GRUAN data products can voluntarily register their use of the data. This would:

- Allow the Lead Centre to maintain statistics on data usage. This would be useful when applying for funding to support GRUAN operations.
- Allow users of data to be informed if and when newer versions of the data become available.
- Facilitate reporting of potential errors/anomalies in the data by end-users.

[11] http://nco.sourceforge.net/

Such a facility will need to exist independently of the GRUAN NCDC archives to avoid legal issues related to data retention by US government agencies.

As discussed above, GRUAN sites are likely to also be members of other networks and are likely to submit data to end-users through other network's archives. Data submitted through non-GRUAN networks may be subject to different data processing, different QA/QC procedures, and different calibrations resulting in a data product that is different from the GRUAN product. This is not seen as a problem since the product delivered through other networks is not identified as 'GRUAN' data.

Users of GRUAN data need to know the version of any dataset they are using and whether newer versions might be available. The names of data files must therefore include the data version identifier to facilitate easy identification of the data version. An application to periodically check for updates of GRUAN data files found on a client computer with the database currently at NCDC needs to be developed.

8.7 Data archiving

GRUAN does not necessarily need to build its own data archive and user interface. This is a rather costly operation for any large network and partnering with an established data archive such as NCDC with a user-friendly interface is preferred. Because data cannot be quality assured or corrected in near real-time, additional processing steps and uncertainty estimate assignment will be required. This key processing will be allowed to grow, and thus, data versioning will be required. It is important that the GRUAN archive includes all previous versions of any given data set so that analyses using previous versions of data can be repeated if required.

8.8 Quality control at the instrument/site level

Part of the data management within GRUAN includes feedback to the sites in the form of reports on data submission, data quality, and comprehensiveness of metadata submitted. Existing algorithms, potentially supplemented by future algorithms to be developed, shall be used operationally to identify systematic errors, anomalies or instrumental issues. Results of such tests shall be communicated back to GRUAN sites on short timescales so that remedial action can be taken if required. Following the example of the ARM Data Quality Office[12], communicating quality control results to GRUAN site operators and engineers will facilitate improved instrument performance and thereby minimize the amount of unacceptable data collected.

[12] http://dq.arm.gov/

9 POST-PROCESSING ANALYSIS AND FEEDBACK

Analysis of GRUAN data products by end-users will need to be sensitive to data versioning. As new knowledge becomes available and data are reprocessed as a result, newer versions of data sets will be provided through the GRUAN archives and end-users need to be aware of such updates and, if necessary, repeat their own analyses. Users of GRUAN data must always document the version of data used to ensure that the analyses can be independently replicated. Key to this process will be the ability to make users aware of updated versions of data sets that they previously accessed, now becoming available. The data processing centre, either the Lead Centre or the designated GRUAN site/facility specializing in processing of that particular data set, should be tasked with data version control and ensuring that the necessary metadata on data versions are made available to end-users.

Inevitably, algorithms change and errors in data processing occur that are not necessarily apparent until the data are used. Therefore, a facility that allows data users to report potential bugs or anomalies found in data during analyses of the data needs to be designed and implemented. This might be modelled on the ARM Program Climate Research Facility bug reporting system. This is as yet not mature but will be fully documented once developed.

A quality system should include procedures for feeding back into the measurement and quality control process to prevent the errors from recurring. Quality assurance can be applied in real-time post measurement, and can feed into the quality control process for the next process of a quality system, but in general it tends to operate in non-real time. Again, it will be essential to document how this is done once the process is sufficiently mature.

10 QUALITY MANAGEMENT

This chapter defines the principles and the methodological framework for GRUAN operations, and details how activities will be coordinated to manage and control data quality within GRUAN. Quality management within GRUAN consists of quality assurance and quality control.

Quality assurance (QA): The purpose of quality assurance is to provide confidence that the requirements for achieving quality will be fulfilled. QA includes all the planned and systematic activities that will be implemented such that quality requirements for a product or service will be fulfilled.

Quality control (QC): The purpose of quality control is to ensure that the expectations created by QA are fulfilled. QC is associated with those operational methods, techniques and activities used to ensure that the quality requirements (as defined by QA) are fulfilled.

The GRUAN quality management policy is to achieve a level of data quality that allows the primary goals of GRUAN (see Section 1.2) to be met for all potential users of GRUAN data products. Assuring the quality of the GRUAN data begins with a robust process of describing, quantifying and validating all sources of uncertainty in all GRUAN measurements and by providing rich metadata that describe all facets of the measurement process. Where total measurement uncertainties are below some prescribed threshold, this increases confidence in the quality of the GRUAN data. The use of redundant measurements, as described in Section 3.1.3, also serves to assure the quality of the GRUAN data products. Agreement of two independent measurements, preferably based on different measurement principles, provides a high degree of confidence that no significant systematic effect was disregarded and uncertainties were not under-estimated. Laboratory tests and intercomparisons are fundamental methods for establishing and confirming uncertainty estimates for GRUAN data products. Laboratory tests provide an opportunity to investigate in detail the performance of instruments under controlled conditions and to measure differences against certified references or other standards. Data from these experiments can be used to detect biases that may be corrected for and to determine calibration uncertainties. Field intercomparisons allow multiple *in situ* sensors and remote sensing data to be directly compared under the actual atmospheric conditions of the required measurement, including the complex environmental conditions (temperature, humidity, pressure, wind/flow rate, radiation, and chemical composition) that cannot be fully reproduced in the laboratory. These complementary activities increase confidence that measurements are subject to neither unanticipated effects nor undiscovered systematic uncertainties. Therefore field experiments are particularly useful for assuring the quality of GRUAN data products. The use of GRUAN data in meteorological reanalyses also adds to the assurance of GRUAN data quality since the measurements, with their uncertainties, can be tested for comparability with the data assimilation model values in an assimilation setting within the known internal variability of the system.

Quality control will be achieved through the application of the various measurement protocols defined in this Guide and in related measurement system guides. To the extent possible, visual inspection of all data by science/instrument experts will be required for all instruments to minimize anomalies that slip through automated routines. The Lead Centre shall coordinate this effort, which shall be distributed across different GRUAN sites and other interested parties as deemed appropriate including Task Teams and members of WG-GRUAN. Vertically resolved uncertainty estimates, prepared independently for each site, will be used as a metric to compare the site-to-site quality of the observations.

Section 4 of this guide provides explicit requirements regarding random errors, systematic errors, stability, resolution and representativeness for measurements made within GRUAN. Minimizing cost without compromising quality is also an implied or explicit requirement for measurements made within GRUAN. The purpose of quality management is to ensure that GRUAN data meet the requirements in terms of uncertainty, resolution, continuity, homogeneity, representativeness, timeliness, format etc. for their intended use, at a minimum practicable cost. GRUAN recognizes that all measurements are imperfect, but, if their quality is known and demonstrable, they can be used appropriately.

Quality management is required at all points in the measurement process from network planning and training, through installation and site operations to data transmission and archiving. This quality management must include feedback and follow-up provisions across a range of timescales from near real-time to annual reviews. Because of the emphasis on the provision of robust measurement uncertainties and the associated requirement for in-depth quality management, the resources required within GRUAN to undertake quality management will likely be a significant proportion of the cost of operating the network, and very likely more than the few percent of overall operating costs typical of many observational networks. However, without this expenditure, the quality of the data will be unknown, and their usefulness diminished.

A key aspect of quality management within GRUAN will be fulfilling customer requirements. To this end systems shall be developed to:

1. Inform users of GRUAN products of changes in measurements systems at specific sites.
2. Provide an incident reporting system that can flag data anomalies to users.
3. Inform users of the availability of updates to previously accessed data products.
4. Provide ‘help desk’ support to users of GRUAN data products.

Establishing close working relationships with instrument manufacturers will also be central to quality assurance within GRUAN.

A common component of quality assurance is quality monitoring or performance monitoring, a non-real-time activity in which the performance of the network or observing system is examined for trends and systematic deficiencies. Performance monitoring within GRUAN will primarily be the responsibility of the Lead Centre, supported by WG-GRUAN and task teams, but where other specialists may be co-opted to assist in performance assessments. The outcomes of recertification of GRUAN sites (see Section 5.5) and GRUAN site audits (see Section 5.6) will be an essential component of performance monitoring. Requests for external, independent assessments of GRUAN performance from key user groups of GRUAN data products might also serve a useful performance monitoring function. The development of quantitative performance indicators such as:

1. Data downloads,
2. The number of peer reviewed publications in which GRUAN data have been used,
3. Scientific case studies of the added value resulting from the use of GRUAN data products,
4. The number of GRUAN projects funded through national or international funding agencies.

may serve to provide year-to-year traceability of GRUAN’s impact within the climate community.

ACRONYMS

AERONET: AErosol RObotic NETwork
AIRS: Atmospheric InfraRed Sounder
ARM: Atmospheric Radiation Measurement programme
ACRF: ARM Program Climate Research Facility
AOD: Aerosol Optical Depth
AOPC: Atmospheric Observation Panel for Climate
AWI: Alfred Wegener Institute
BSRN: Baseline Surface Radiation Network
CBS: WMO Commission for Basic Systems
CDR: Climate Data Record
CFH: Cryogenic Frostpoint Hygrometer
CIMO: WMO Commission for Instruments and Methods of Observation
CRD: Converted Raw Data
DIAL: DIfferential Absorption Lidars
DQ: Data Quality
DWD: German Meteorological Service
EARLINET: European Aerosol Research Lidar Network
ECV: Essential Climate Variable
FLASH-B: Fluorescent Advanced Stratospheric Hygrometer for Balloon
FPH: Frost-Point Hygrometer
FTIR: Fourier-Transform InfraRed
FTS: Fourier-Transform Spectrometers
GALION: GAW Atmospheric LIdar Observation Network
GATNDOR: GRUAN Analysis Team for Network Design and Operations Research
GAW: Global Atmosphere Watch
GCOS: Global Climate Observing System
GEWEX: Global Energy and Water Cycle EXperiment
GHG: Well-mixed greenhouse gas (CO_2, CH_4, N_2O, CFCs, HFCs, PFCs, SF_6, etc.)
GLASS: GRUAN Lidar Analysis Software System
GNSS: Global Navigation Satellite System
GOS: Global Observing System
GPS: Global Positioning System
GPS-RO: Global Positioning System Radio Occultation
GRUAN: GCOS Reference Upper Air Network
GSICS: Global Space-based InterCalibration System
GTS: Global Telecommunication System
GUAN: GCOS upper-air network
HITRAN: HIgh-resolution TRANsmission molecular absorption database
HMEI: Hydro-Meteorological Equipment Industry
ICM: Implementation - Coordination Meeting (GRUAN)

IGLIMP: Individual GRUAN Lidar Instrumentation and Measurement Protocol
IGPD: Integrated GRUAN Product Data
ILW: Integrated Liquid Water
ISCCP: International Satellite Cloud Climatology Project
IWV: Integrated Water Vapour
LST: Local Solar Time
MWR: Microwave Radiometers
NCDC: NOAA National Climate Data Centre
NDACC: Network for the Detection of Atmospheric Composition Change
NMS: National Meteorological Service
NMHS: National Hydrological and Hydrometeorological Services
NMI: National Metrological Institute
NOAA: National Oceanic and Atmospheric Administration
NRT: Near real time (within 2 hours of a measurement)
NRTD: Near real time data
NWP: Numerical Weather Prediction
OLR: Out-going Longwave Radiation
PDF: Probability Density Function
PRD: Primary Raw Data
PMOD: Physikalisch-Meteorologisches Observatorium Davos
RMS: Root Mean Square
PR: Permanent Representative (of WMO to a member country)
SASBE: Site Atmospheric State Best Estimate
SCOPE-CM: Sustained, Co-Ordinated Processing of Environmental Satellite Data for Climate Monitoring
SGPD: Standard GRUAN Product Data
SHADOZ: Southern Hemisphere Additional Ozonesondes
SNR: Signal to Noise Ratio
SRB: Surface Radiation Budget
TCCON: Total Carbon Column Observing Network
UT/LS: Upper troposphere/lower stratosphere
WCDMP: World Climate Data and Monitoring Programme
WDAC: WCRP Data Advisory Council
WG-GRUAN: Working Group on GRUAN
WIS: WMO Information System
WMO: World Meteorological Organisation
WOUDC: World Ozone and UV Data Centre
WRC: World Radiation Centre
WWW: World Weather Watch

Appendix A – Expanded details on additional GRUAN Essential Climate Variables

Appendix 1 of GCOS-134 details the measurement requirements for these variables. This section does not repeat those requirements but provides commentary on those stated requirements.

A.1. Wind speed (priority 2)

The high accuracy of 0.5 m/s prescribed for wind speed is needed to delineate calm conditions from light winds.

A.2. Wind direction (priority 2)

No supplementary comments yet.

A.3. Ozone (priority 2)

During a discussion at the ICM-2 meeting, it was suggested that ozone should develop into a priority 1 variable for GRUAN. The consensus appears to be that it remains a priority 2 variable.

A.4. Methane (priority 2)

No supplementary comments yet.

A.5. Net radiation (priority 2)

The prescribed precision and accuracy values of 5 W/m^2 match the requirements for the BSRN network.

A.6. Shortwave downward radiation (priority 2)

The stated measurement range of 0 to 2000 W/m^2 exceeds the solar constant (1366 W/m^2) but is required since in the presence of partly cloudy skies and when the sun is not obscured by cloud, reflections off clouds can enhance surface shortwave radiation significantly. The prescribed precision and accuracy values of 3 and 5 W/m^2 respectively, match the requirements for the BSRN network.

A.7. Shortwave upward radiation (priority 2)

The prescribed precision of 2 W/m^2 and accuracy of 3% match the requirements for the BSRN network.

A.8. Longwave downward radiation (priority 2)

The prescribed precision and accuracy values of 1 and 3 W/m^2 respectively, match the requirements for the BSRN network.

A.9. Longwave upward radiation (priority 2)

The prescribed precision and accuracy values of 1 and 3 W/m^2 respectively, match the requirements for the BSRN network.

A.10. Radiances (priority 2)

The stated stability requirement of 0.03%/decade is achievable through SI traceability. The precision and accuracy requirements of 0.01% and 0.15% respectively are applicable for mean seasonal radiances at ~1000 km spatial scale.

A.11. Aerosol optical depth (priority 2)

Measurements of all aerosol parameters should be spectrally resolved. The aerosol optical depth is the most important of the aerosol parameters. While the other aerosol parameters will be scientifically useful if the aerosol optical depth is large, when the aerosol optical depth is small, measurements of other aerosol parameters become less valuable.

A.12. Aerosol total mass concentration (priority 2)

Size-fractionated measurements are required.

A.13. Aerosol chemical mass concentration (priority 2)

Size-fractionated measurements are required.

A.14. Aerosol light scattering (priority 2)

Size-fractionated and spectral measurements are required.

A.15. Aerosol light absorption (priority 2)

Size-fractionated and spectral measurements are required.

A.16. Cloud amount/frequency (priority 2)

The prescribed precision and accuracy ranges of 0.1%-0.3% result from cloud variations of 1-3% found in the ISCCP database. The prescribed long-term stability requirement of 0.1%-0.2% results from the 1-2%/decade trends found by Norris (2005).

A.17. Cloud base height (priority 2)

The prescribed measurement range of 0-20 km (1000-50 hPa) is consistent with the vertical cloud range found in Rossow and Schiffer (1999). The prescribed precision and accuracy of 100 m (10-40 hPa) is consistent with variations derived from the ISCCP database. The long-term stability requirement of 20m/decade is what would be required to detect the trend in global mean cloud base height of 44 m/decade reported by Chernykh et al. (2001)[13].

A.18. Cloud layer heights and thicknesses (priority 2)

The prescribed vertical resolution of 50 m is required to resolve cloud layer thickness of ~30 m for cirrus clouds and is easily achievable with a lidar based system (Winker and Vaughan, 1994).

A.19. Carbon Dioxide (priority 3)

This ECV was not included in Appendix 1 of GCOS-112 but is key to understanding trends in tropospheric stratospheric temperatures and so is included here.

A.20. Cloud top height (priority 3)

Cloud top height measurements are also important for radiosonde temperature uncertainty analysis. When a radiosonde emerges into dryer air above a cloud ,evaporation of the condensed water cools the sensor and creates a cool bias in this region. This effect can lead to deviations up to 1K above a cloud and the data need to be flagged appropriately, e.g., by assigning a correspondingly increased uncertainty to data in such regions.

[13] Trends reported in Chernykh have been questioned by Seidel and Durre (2003)

A.21. Cloud top pressure (priority 3)

No supplementary comments yet.

A.22. Cloud top temperature (priority 3)

No supplementary comments yet.

A.23. Cloud particle size (priority 4)

No supplementary comments yet.

A.24. Cloud optical depth (priority 4)

No supplementary comments yet.

A.25. Cloud liquid water/ice (priority 4)

No supplementary comments yet.

REFERENCES

Austin, J., J. Wilson, F. Li and H. Vömel, Evolution of Water Vapor Concentrations and Stratospheric Age of Air in Coupled Chemistry-Climate Model Simulations, *J. Atmos. Sci.*, 64, 905-921, 2007.

Bennett, L.J., T.M. Weckwerth, A.M. Blyth, B. Geerts, Q. Miao and Y.P. Richardson, Observations of the evolution of the nocturnal and convective boundary layers and the structure of open-celled convection on 14 June 2002, *Monthly Weather Review*, 138, 2589, 2010.

Blackwell, W.J. and F.W. Chen, Neural Networks in Atmospheric remote Sensing, Artech House, 2009.

Bodeker, G.E. and R.L. McKenzie, An algorithm for inferring surface UV irradiance including cloud effects, *J. Appl. Meteorol.*, 35(10), 1860-1877, 1996.

Bodeker, G.E., I.S. Boyd and W.A. Matthews, Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986-1996, *J. Geophys. Res.*, 103(D22), 28661-28681, 1998.

Boers, R. and E. van Meijgaard, What are the demands on an observational program to detect trends in upper tropospheric water vapor anticipated in the 21^{st} century?, *Geophys. Res. Lett.*, 36, L19806, doi:10.1029/2009GL040044, 2009.

Boyd, I.S., G.E. Bodeker, B.J. Connor, D.P.J. Swart and E.J. Brinksma, An assessment of ECC ozonesondes operated using 1% and 0.5% KI cathode solutions at Lauder, New Zealand, *Geophys. Res. Lett.*, 25(13), 2409-2412, 1998.

Brinksma, E.J., J.B. Bergwerff, G.E. Bodeker, K.F. Boersma, I.S. Boyd, B.J. Connor, J.F. de Haan, W. Hogervorst, J.W. Hovenier, A. Parrish, J.J. Tsou, J.M. Zawodny and D.P.J. Swart, Validation of 3 years of ozone measurements over Network for the Detection of Stratospheric Change station Lauder, New Zealand, *J. Geophys. Res.*, 105(D13), 17291-17306, 2000.

Chernykh, I.V., O.A. Alduchov and R.E. Eskridge, Trends in low and high cloud boundaries and errors in height determination of cloud boundaries, *Bull. Amer. Meteor. Soc.*, 82(9), 1941-1947, 2001.

Dai, A., J. Wang, P.W. Thorne, D.E. Parker, L. Haimberger and X.L. Wang, A new approach to homogenize daily radiosonde humidity data, *J. Climate*, 24, 965-991, 2011.

Daumont, D., J. Brion, J. Charbonnier and J. Malicet, Ozone UV Spectroscopy I: Absorption Cross-Sections at Room Temperature, *J. Atmos. Chem.*, 15, 145-155, 1992.

Demoz, B., C. Flamant, T. Weckwerth, D. Whiteman, K. Evans, F. Fabry, P. Di Girolamo, D. Miller, B. Geerts, W. Brown, G. Schwemmer, B. Gentry, W. Feltz and Z. Wang, The Dryline on 22 May 2002 during IHOP-2002: Convective-Scale Measurements at the Profiling Site, *Monthly Weather Review*, 134, 294–310, 2006.

Ferrare, R.A., E.V. Browell, S. Ismail, S.A. Kooi, L.H. Brasseur, V.G. Brackett, M.B. Clayton, J.D. Barrick, G.S. Diskin, J.E.M. Goldsmith, B.M. Lesht, J.R. Podolske, G.W. Sachse, F.J. Schmidlin, D.D. Turner, D.N. Whiteman, D. Tobin, L.M. Miloshevich, H.E. Revercomb, B.B Demoz and P. Di Girolamo, Characterization of Upper-Troposphere Water Vapor Measurements during AFWEX Using LASE, *J. Atmos. Ocean. Tech.*, 21, 1790-1808, 2004.

Forster, P.M., G.E. Bodeker, R. Schofield, S. Solomon and D.W.J. Thompson, Effects of ozone cooling in the tropical lower stratosphere and upper troposphere, *Geophys. Res. Lett.*, 34, L23813, doi:23810.21029/22007GL031994, 2007.

Fueglistaler, S. and P.H. Haynes, Control of interannual and longer-term variability of stratospheric water vapor, *J. Geophys. Res.*, 110, D24108, doi:24110.21029/22005JD006019, 2005.

GCOS-92, Implementation plan for the global observing system for climate in support of the UNFCCC.WMO/TD No. 1219, 2004.

GCOS-112, GCOS Reference Upper-air Network (GRUAN): Justification, requirements, siting and instrumentation options. WMO/TD-No.1379, 2007.

GCOS-121, GCOS Reference Upper-air Network (GRUAN): Report of the GRUAN Implementation Meeting, Lindenberg, Germany, 26-28 February 2008.

GCOS-122, Fourteenth Session of the GCOS/WCRP Atmospheric Observation Panel for Climate (AOPC-XIV), Geneva, 12-25 April 2008, WMO/TD No. 1436, 2008.

GCOS-131, Report of the first GCOS reference upper-air network implementation and coordination meeting (GRUAN ICM-1). WMO/TD No. 1492, 2009.

GCOS-134, GRUAN Implementation Plan 2009-2013. WMO/TD No. 1506, 2009.

GCOS-138, Implementation Plan for the Global Observing System for Climate in support of the UNFCCC (2010 Update). WMO-TD/No. 1523, 2010.

GCOS-140, Report of the Second GCOS Reference Upper-air Network Implementation and Coordination Meeting (GRUAN ICM-2), WMO/TD No. 1526, 2009.

GCOS-143, Guideline for the Generation of Datasets and Products Meeting GCOS Requirements, WMO/TD No. 1530, 2010.

GCOS-148, Summary Report and Recommendations from the Sixteenth Session of the GCOS/WCRP Atmospheric Observation Panel for Climate (AOPC-XVI). WMO/TD No. 1574, 2011.

GCOS-155, WIGOS Pilot Project Meeting on GRUAN Observing Practices and Governance, 2012.

Gettelman, A., W.J. Randel, F. Wu and S.T. Massie, Transport of water vapor in the tropical tropopause layer, *Geophys. Res. Lett.*, 29, 10.1029/2001GL013818, 2002.

Haimberger, L., Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics, *J. Climate*, 20, 1377-1403, 2007.

Hurst, D.F., E.G. Hall, A.F. Jordan, L.M. Miloshevich, D.N. Whiteman, T. Leblanc, D. Walsh, H. Vömel and S.J. Oltmans, Comparisons of temperature, pressure and humidity measurements by balloon-borne radiosondes and frost point hygrometers during MOHAVE-2009, *Atmos. Meas. Tech.*, 4, 2777–2793, doi:10.5194/amt-4-2777-2011, 2011a.

Hurst, D.F., S.J Oltmans, H. Vömel, K.H. Rosenlof, S.M. Davis, E.A. Ray, E.G. Hall and A.F. Jordan, Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record, *J. Geophys. Res.*, 116, D02306, doi:10.01029/2010JD015065, 2011b.

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

IOM report-No. 107/WMO/TD-No. 1580, WMO intercomparison of high quality radiosonde systems, 2011.

Koch, S.E., W. Feltz, F. Fabry, M. Pagowski, B. Geerts, K.M. Bedka, D.O. Miller and J.W. Wilson, Turbulent mixing processes in atmospheric bores and solitary waves deduced from profiling systems and numerical simulation, *Monthly Weather Review*, 136, doi:10.1175/2007MWR2252.1, 2008.

Leblanc, T., T.D. Walsh, I.S. McDermid, G.C. Toon, J.-F. Blavier, B. Haines, W.G. Read, B. Herman, E. Fetzer, S. Sander, T. Pongetti, D.N. Whiteman, T.G. McGee, L. Twigg, G. Sumnicht, D. Venable, M. Calhoun, A. Dirisu, D. Hurst, A. Jordan, E. Hall, L. Miloshevich, H. Vömel, C. Straub, N. Kampfer, G.E. Nedoluha, R.M. Gomez, K. Holub, S. Gutman, J. Braun, T Vanhove, G. Stiller and A. Hauchecorne, Measurements of Humidity in the Atmosphere and Validation Experiments (MOHAVE)-2009: overview of campaign operations and results, *Atmos. Meas. Tech.*, 4, 2579–2605, 2011.

Leiterer, U., H. Dier and T. Naebert, Improvements in radiosonde humidity profiles using RS80/RS90 radiosondes of Vaisala, *Contrib. Atmos. Phys.*, 70, 319–336, 1997.

McDermid, I.S., J.B. Bergwerff, G.E. Bodeker, I.S. Boyd, E.J. Brinksma, B.J. Connor, R. Farmer, M.R. Gross, P. Kimvilakani, W.A. Matthews, T.J. McGee, F.T. Ormel, A. Parrish, U. Singh, D.P.J. Swart and J.J. Tsou, OPAL: Network for the detection of stratospheric change ozone profiler assessment at Lauder, New Zealand 2. Intercomparison of revised results, *J. Geophys. Res.*, 103(D22), 28693-28699, 1998.

McDermid, I.S., J.B. Bergwerff, G.E. Bodeker, I.S. Boyd, E.J. Brinksma, B.J. Connor, R. Farmer, M.R. Gross, P. Kimvilakani, W.A. Matthews, T.J. McGee, F.T. Ormel, A. Parrish, U. Singh, D.P.J. Swart, J.J. Tsou, P.H. Wang and J. Zawodny, OPAL: Network for the detection of stratospheric change ozone profiler assessment at Lauder, New Zealand 1. Blind intercomparison, *J. Geophys. Res*., 103(D22), 28683-28692, 1998.

Melfi, S.H., D. Whiteman and R. Ferrare, Observation of atmospheric fronts using raman lidar moisture measurements, *J. Appl. Meteor.*, 28, 789, 1989.

Norris, J.R., Multidecadal changes in near-global cloud cover and estimated cloud cover radiative forcing, *J. Geophys. Res.*, 110, D08206, doi:08210.01029/02004JD005600, 2005.

Ohring, G., B. Wielicki, R. Spencer, B. Emery and R. Datla, Satellite Instrument Calibration for Measuring Global Climate Change: Report of a Workshop, *Bull. Am. Meteorol. Soc.*, 86, 1303-1313, 2005.

Peter, T., C. Marcolli, P. Spichtinger, T. Corti, M.B. Baker and T. Koop, When Dry Air Is Too Humid, *Science*, 314, 1399–1402, DOI: 10.1126/science.1135199, 2006.

Po-Chedley, S. and Q. Fu, A bias in the mid-tropospheric channel warm target factor on the NOAA-9 microwave sounding unit, *J. Atmos. Oceanic Tech.*, doi:10.1175/JTECH-D-11-00147.1, 2012.

Reale, A.L., B. Sun, M. Pettey and F.W. Tilley, The NOAA Products Validation System (NPROVS), JTECH-A, 29 (5), 629-645, 2012

Rodgers, C.D. Inverse Methods For Atmospheric Sounding: Theory and Practice, World Scientific Publishing Co. Pte. Ltd., 2, 240, 2000.

Rosenlof, K.H., S.J. Oltmans, D. Kley, J.M. Russell III, E.-W. Chiou, W.P. Chu, D.G. Johnson, K.K. Kelly, H.A. Michelsen, G.E. Nedoluha, E.E. Remsberg, G.C. Toon and M.P. McCormick, Stratospheric water vapour increases over the past half-century, *Geophys. Res. Lett.*, 28 (7), 1195-1198, 2001.

Rossow, W.B. and R.A. Schiffer, Advances in Understanding Clouds from ISCCP, *Bull. Amer. Meteor. Soc.*, 80 (11), 2261-2287, 1999.

Schmidlin, F.J., Derivation and application of temperature corrections for the United States radiosonde, in *Symposium on Meteorological Observations and Instrumentations*, 7th, New Orleans, LA, 14–18 January 1991, Preprints (A92-32051 12-47), Boston, MA, American Meteorological Society, 227–231, 1991.

Schubert, S., D. Dee, S. Uppala, J. Woollen, J. Bates and S. Worley, Report of the Workshop on "The Development of Improved Observational Data Sets for Reanalysis: Lessons learned and Future Directions", 31 pp., http://gmao.gsfc.nasa.gov/pubs/docs/Schubert273.doc, 2006.

Seidel, D.J. and M. Free, Measurement Requirements for Climate Monitoring of Upper-air Temperature Derived from Reanalysis Data, *J. Climate*, 19, 854-871, 2006.

Seidel, D.J., F.H. Berger, H.J. Diamond, J. Dykema, D. Goodrich, F. Immler, W. Murray, T. Peterson, D. Sisterson, M. Sommer, P. Thorne, H. Vömel and J. Wang, Reference upper-air observations for climate: rationale, progress, and plans, *Bull. Amer. Meteor. Soc.*, 3, 361-369, 2009.

Seidel, D.J. and I. Durre, Comments on "Trends in low and high cloud boundaries and errors in height determination of cloud boundaries", *Bull. Amer. Meteor. Soc.*, 84(2), 237-240, 2003.

Seidel, D.J., F.H. Berger, H.J. Diamond, J. Dykema, D. Goodrich, F. Immler, W. Murray, T. Peterson, D. Sisterson, M. Sommer, P. Thorne, H. Vömel, and J. Wang, Reference upper-air observations for climate: rationale, progress, and plans, *Bull. Amer. Meteor. Soc.*, 3, 361-369, 2009.

Seidel, D.J., N.P. Gillett, J.R. Lanzante, K.P. Shine and P.W. Thorne, Stratospheric temperature trends: our evolving understanding, *WIREs Climate Change*, 2, 592-616, 2010.

Sherwood, S.C. and Dessler, A.E., A model for transport across the tropical tropopause, *J. Atmos. Sci.*, 58, 765-779, 2001.

Shine, K.P., M.S. Bourqui, P.M.D. Forster, S.H.E. Hare, U. Langematz, P. Braesicke, V. Grewe, M. Ponater, C. Schnadt, C.A. Smith, J.D. Haigh, J. Austin, N. Butchart, D.T. Shindell, W.J. Randel, T. Nagashima, R.W. Portmann, S. Solomon, D.J. Seidel, J. Lanzante, S. Klein, V. Ramaswamy and M.D. Schwarzkopf, A comparison of model-simulated trends in stratospheric temperatures, *Quarterly Journal of the Royal Meteorological Society*, 129(590), 1565–1588, 2003.

Smit, H.G.J., W. Straeter, B.J. Johnson, S.J. Oltmans, J. Davies, D.W. Tarasick, B. Hoegger, R. Stubi, F.J. Schmidlin, T. Northam, A.M. Thompson, J.C. Witte, I. Boyd and F. Posny, Assessment of the performance of ECC-ozonesondes under quasi-flight conditions in the environmental simulation chamber: Insights from the Juelich Ozone Sonde Intercomparison Experiment (JOSIE), *J. Geophys. Res*., 112, D19306, doi:10.1029/2006JD007308, 2007.

Soden, B.J., D.D. Turner, B.M. Lesht and L.M. Miloshevich, An analysis of satellite, radiosonde, and lidar observations of upper tropospheric water vapor from the Atmospheric Radiation Measurement Program, *J. Geophys. Res.*, 109, D04105, doi:10.1029/2003JD003828, 2004.

Solomon, S., K.H. Rosenlof, R.W. Portmann, J.S. Daniel, S.M. Davis, T.J. Sanford and G.-K. Plattner, Contributions of Stratospheric Water Vapor to Decadal Changes in the Rate of Global Warming, *Science*, 327, 1219-1223, 2010.

Stolarski, R.S. and S.M. Frith, Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty, *Atmos. Chem. Phys., 6*, 4057–4065, 2006.

Suortti, T.M., A. Kats, R. Kivi, N. Kämpfer, U. Leiterer, L.M. Miloshevich, R. Neuber, A. Paukkunen, P. Ruppert, H. Vömel and V. Yushkov, Tropospheric Comparisons of Vaisala Radiosondes and Balloon-Borne Frost-Point and Lyman-α Hygrometers during the LAUTLOS-WAVVAP Experiment, *J. Atmosph. Oceanic Tech*., 25, 149-166, 2008.

Thompson, A.M., J.C. Witte, H.G.J. Smit, S.J. Oltmans, B.J. Johnson, V.W.J.H. Kirchhoff and F.J. Schmidlin, Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2004 tropical ozone climatology: 3. Instrumentation, station-to-station variability, and evaluation with simulated flight profiles, *J. Geophys. Res.*, 112, D03304, doi:10.1029/2005JD007042, 2007.

Thorne, P.W., D.E. Parker, J.R. Christy and C.A. Mears, Uncertainties in climate trends: Lessons from upper-air temperature records, *Bull. Amer. Meteor. Soc.*, 86, 1437-1442, 2005.

Thorne, P.W., J.R. Lanzante, T.C. Peterson, D.J. Seidel and K.P. Shine, Tropospheric temperature trends: history of an ongoing controversy, *WIREs Climate Change*, 2, 66-88, 2010.

Tiao, G.C., G.C. Reinsel, D. Xu, J.H. Pedrick, X. Zhu, A.J. Miller, J.J. DeLuisi, C.L. Mateer and D.J. Wuebbles, Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, *J. Geophys. Res.*, 95(D12), 20507-20517, 1990.

Tobin, D.C., H.E. Revercomb, R.O. Knuteson, B.M. Lesht, L.L. Strow, S.E. Hannon, W.F. Feltz, L.A. Moy, E.J. Fetzer and T.S. Cress, Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, *J. Geophys. Res.*, 111, D09S14, doi:10.1029/2005JD006103, 2006.

Trenberth, The need for a new upper-air baseline network, [details from Kevin to come in here], 2003.

Vömel, H., J.E. Barnes, R.N. Forno, M. Fujiwara, F. Hasebe, S. Iwasaki, R. Kivi, N. Komala, E. Kyrö, T. Leblanc, B. Morel, S.-Y. Ogino, W.G. Read, S.C. Ryan, S. Saraspriya, H. Selkirk, M. Shiotani, J.V. Canossa and D.N. Whiteman, Validation of Aura Microwave Limb Sounder water vapor by balloonborne Cryogenic Frost point Hygrometer measurements, *J. Geophys. Res.*, 112, D24S37, doi:10.1029/2007JD008698, 2007a.

Vömel, H., D.E. David and K. Smith, Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations, *J. Geophys. Res.*, 112, D08305, doi:10.1029/2006JD007224, 2007b.

Wang, J., D.J. Carlson, D.B. Parsons, T.F. Hock, D. Lauritsen, H.L. Cole, K. Beierle and E. Chamberlain, Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication, *Geophys. Res. Lett.*, 30, 1860, doi:10.1029/2003GL016985, 2003.

Wang, J., and L. Zhang, Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements, *J. Climate*, 21, 2218-2238, 2008.

Wang, J.S., D.J. Seidel and M. Free. How well do we know recent climate trends at the tropical tropopause? *J. Geophys. Res.*, 117, D09118, doi:10.1029/2012JD017444, 2012

Weatherhead, E.C., G.C. Reinsel, G.C. Tiao, X.-L. Meng, D. Choi, W.-K. Cheang, T. Keller, J. DeLuisi, D.J. Wuebbles, J.B. Kerr, A.J. Miller, S.J. Oltmans and J.E. Frederick, Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, 103(D14), 17149-17161, 1998.

Whiteman, D.N., K.C. Vermeesch, L.D. Oman and E.C. Weatherhead, The relative importance of random error and observation frequency in detecting trends in upper tropospheric water vapor, *J. Geophys. Res.*, 116, D21118, doi:10.1029/2011JD016610, 2011.

Winker, D.M. and M.A. Vaughan, Vertical distribution of clouds over Hampton, Virginia observed by lidar under the ECLIPS and FIRE ETO programs, *Atmos. Res.*, 34, 117-133, 1994.

WMO-No. 8, Guide to Meteorological Instruments and Methods of Observation, 7th edition, 2008.

WMO-No. 488, Guide to the Global Observing System, 2010.

WMO-No. 100, Guide to Climatological Practices, 2011.

World Meteorological Organization, WMO Intercomparison of High Quality Radiosonde Systems, Yangjiang, China (J. Nash, T. Oakley, H. Vömel and Li Wei) Instruments and Observing Methods Report No. 107, WMO/TD-No 1580, Geneva, 2011.

Zhou, S., M.A. Geller and M. Zhang, Cooling trend of the tropical cold point tropopause temperatures and its implications, *J. Geophys.Res.*, 106, 1511-1522, 2001.